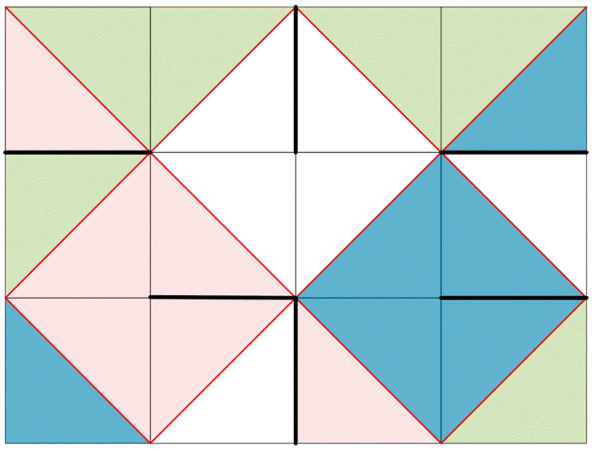
A MATHEMATICAL TRIBUTE

to Professor

José María Montesinos Amilibia



DEPARTAMENTO DE GEOMETRÍA Y TOPOLOGÍA
FACULTAD DE CIENCIAS MATEMÁTICAS – UCM

A MATHEMATICAL TRIBUTE
to Professor José María Montesinos Amilibia

Edita
Departamento de Geometría y Topología
Facultad de Ciencias Matemáticas – UCM
Plaza de Ciencias, 3
Ciudad Universitaria
28040 Madrid

Cubierta
Raquel Díaz
Tetraedro inscrito en un cubo
(Dóblese por las líneas negras finas, córtese por las líneas negras gruesas.)

ISBN: 978-84-608-1684-3
Depósito Legal: M-2056-2016

Imprenta
Ulzama Digital

Impreso en España

Professor José María Montesinos Amilibia

Oh my knots!

This volume contains the contributions presented by several colleagues as a tribute to the mathematical and human qualities of José María Montesinos Amilibia on the occasion of his seventieth birthday. The editors would like to express their thanks to the contributors and their very especial gratitude to José María for his example through many years of scientific and personal contact.

Marco Castrillón
Elena Martín-Peinador
José M. Rodríguez-Sanjurjo
Jesús M. Ruiz

Muchas gracias, José María

Este es un acto importante para los que estamos aquí, algunos de los cuales conocen a José María Montesinos desde hace más de 50 años. Hemos venido para darle las gracias por muchas cosas, pero sobre todo por su ejemplo, el ejemplo de cómo es posible hacer cosas valiosas en tiempo difíciles o muy difíciles, como fueron los que vivió en la primera etapa de su investigación, desde 1967 hasta que se fue a Estados Unidos en 1976. Aquí, en la facultad, solo y por sus propios medios, realizó los que, según su propio testimonio, son sus mejores trabajos. Esto se puede constatar también en su página web de la Academia donde cita los que cree son sus trabajos más interesantes, todos pertenecientes a aquella época. Algunos publicados en revistas muy relevantes, como el Bulletin of the AMS, otros en revistas más modestas, pero todos conteniendo resultados excelentes. También la crítica, en particular el Mathematical Reviews, ha sido explícita en su reconocimiento de la calidad de los trabajos de aquel periodo. Posteriormente Montesinos conoció y aprendió mucho de Thurston y de otros: Casson, Kirby, Matsumoto, Edwards, Siebenmann y Fico González Acuña, y siempre ha manifestado su admiración y reconocimiento hacia ellos. Evidentemente, su actividad científica posterior se ha beneficiado enormemente de estos contactos y de colaboraciones como la que ha mantenido con María Teresa Lozano y Mike Hilden a lo largo de tantos años.

En sentido inverso se puede decir que otros aprendieron no menos de él. Esto se ha podido comprobar en el flujo continuo de visitantes de todo el mundo que han venido al departamento para aprender. Éste es el sentido opuesto al que nos ha llevado a la mayoría a ir fuera para adquirir conocimiento.

Montesinos ha obtenido un gran reconocimiento científico. Él ha valorado y agradecido el reconocimiento, científico y humano, de sus colegas. Especialmente importante para él fue el que le dio Ralph Fox en su etapa inicial. Gracias al apoyo que le prestó, pudo saber que iba en la buena dirección y que sus resultados eran significativos. Desgraciadamente, Fox falleció antes del primer viaje de José María a Estados Unidos y no pudo conocerle personalmente. Sin embargo Montesinos no ha buscado los focos, no se siente cómodo cuando es objeto de la atención pública (espero que este acto sea una excepción). En el año 1992 se celebró en París el primer congreso de la Sociedad Matemática Europea. Era una ocasión importante, el lanzamiento de esta sociedad, y fueron invitados diez conferenciantes plenarios de la talla de Arnold, Donaldson y Mumford. Uno de ellos era Montesinos, quizá algunos recuerden los carteles que anunciaban el congreso, en los que él aparecía en esa lista. Sin embargo surgió un problema de financiación y José María renunció a la invitación. Aparte de las razones económicas, creo percibir que esa decisión estuvo motivada por una humildad básica que siempre me ha parecido ver en su personalidad.

Montesinos disfruta del contacto humano, y los demás perciben cuánto pueden ganar estando a su lado y hablando con él. En cualquier reunión científica se puede advertir su popularidad y la relación amistosa de que disfruta con mucha gente. Pero, al mismo tiempo, necesita de la soledad, el retiro. Para su vida interior y para sus teoremas. Los teoremas que fueron hechos en su juventud y los que sigue haciendo en las montañas de Guadarrama y de Gredos. Recientemente nos decía en un seminario que uno de sus últimos resultados le había costado cinco salidas al monte. Le deseo muchas más salidas al monte, muchos más buenos teoremas y que siga viniendo por aquí para contárnoslos. Muchas gracias de nuevo, José María.

Madrid, 8 de septiembre de 2015 *José Manuel Rodríguez-Sanjurjo*
Facultad de Ciencias Matemáticas, UCM

Contents

Morphismes analytiques finis et revêtements ramifiés

Claude Weber

Section de mathématiques. Université de Genève
Claude.Weber@unige.ch

RÉSUMÉ

Le but de cet article est de présenter quelques résultats connus sur les revêtements ramifiés, en topologie et en géométrie analytique complexe. Nous verrons que ces deux aspects sont entremêlés. Les revêtements ramifiés constituent un sujet chargé d'histoire, qui a ses racines dans l'uniformisation des fonctions algébriques. J'essayerai d'en esquisser les grandes lignes.

2010 Mathematics Subject Classification: 32H35, 32C18.

Key words: Variété analytique complexe, revêtement topologique non ramifié, morphisme analytique fini, degré, discriminant.

Notations. En principe, la dimension complexe est notée n. La dimension topologique (disons celle de Lebesgue) est notée m. Le degré d'un revêtement est noté d.

1. Enoncé du théorème principal

Le but de ce texte (rédigé par un topologue paresseux) est d'exposer un théorème connu (mais très utile !) d'existence et d'unicité sur les revêtements ramifiés en géométrie analytique complexe qui, dans ce cas, portent le nom de morphismes finis (finite mappings). Les définitions seront données dans le cours du texte. Autant que possible, on omettra l'adjectif "complexe". Sans plus attendre j'énonce ce résultat sous forme de deux théorèmes. Pris ensemble, ces deux théorèmes constituent le théorème principal.

Théorème 1.1 *Soit Y une variété analytique complexe (lisse par définition de variété analytique). Soit $D \subset Y$ un sous-espace analytique de codimension ≥ 1. Soit*

$$g : Z \longrightarrow Y \setminus D$$

un revêtement topologique non ramifié. (Note sous forme de parenthèse : comme $Y \setminus D$ est lisse et comme le revêtement est non ramifié, on peut munir Z d'une structure de variété analytique de façon à ce que g soit étale (localement un isomorphisme). C'est ce que l'on fait.) Alors il existe :

1) un espace analytique **normal** *X et un sous-espace analytique $B \subset X$ de codimension ≥ 1 tel que $X \setminus B = Z$,*

2) un morphisme analytique fini

$$f : (X, B) \longrightarrow (Y, D)$$

tel que

$$f|\ X \setminus B = Z \longrightarrow Y \setminus D$$

soit égal à g.

Ce premier énoncé peut être vu comme un théorème d'existence, mais aussi de prolongement (on prolonge g en f).

Le deuxième énoncé est un théorème d'unicité. Le point de départ est le même.

Théorème 1.2 *Soient Y ; $D \subset Y$ et $g : Z \longrightarrow Y \setminus D$ comme ci-dessus. On suppose que l'on a un deuxième prolongement $\widetilde{f} : \widetilde{X} \longrightarrow Y$ avec $\widetilde{X}$ normal. Alors il existe un isomorphisme analytique $\varphi : (X, B) \longrightarrow (\widetilde{X}, \widetilde{B})$ tel que $\widetilde{f} \circ \varphi = f$.*

La démonstration est due aux efforts conjugués d'une part de Henri Cartan et de son séminaire et d'autre part de l'école de Heinrich Behnke à Münster : Karl Stein, Hans Grauert et Reinhold Remmert. Elle mêle topologie des revêtements ramifiés et théorie des espaces analytiques. C'est ce que nous allons exposer ci-dessous dans les grandes lignes. Dans le cas des courbes, tout ou partie de ces énoncés s'appelle souvent "le théorème d'existence de Riemann".

2. Morphismes analytiques finis selon Henri Cartan

Tout d'abord il faut savoir ce qu'est un morphisme fini.

Définition 2.1 *Soit $f : X \longrightarrow Y$ un morphisme analytique. Il est appelé* **fini** *s'il est propre et à fibres finies.*

Commentaire. "Propre" est absolument indispensable. L'inclusion d'un ouvert est à fibres finies mais on n'en veut pas. En fait, on verra que très souvent un morphisme fini est ouvert, de sorte qu'il est à la fois ouvert et fermé (puisque propre !) et donc surjectif, si Y est connexe (supposer Y connexe ne coûte rien). Les applications propres entre espaces localement compacts sont étudiées dans [7] p.175. Voir aussi Bourbaki.

Un théorème de Cartan éclaire (mieux illumine) ce qu'il se passe. Voir [4].

Théorème 2.1 *Soient X et Y deux espaces analytiques normaux, connexes de même dimension n. Soit $f : X \longrightarrow Y$ un morphisme analytique fini. Alors :*

1) f est ouverte.

2) Il existe un sous-ensemble analytique $D \subset Y$ de dimension $< n$ et un entier $d \geq 1$ tel que :

i) pour tout $y \in Y$ la fibre $f^{-1}(y)$ se compose d'au plus d points ;

ii) pour tout $y \in Y \setminus D$ la fibre $f^{-1}(y)$ se compose d'exactement d points ;

iii) pour tout $y \in D$ la fibre $f^{-1}(y)$ se compose de moins de d points.

3) L'image réciproque $f^{-1}(D) = B$ est un sous-ensemble analytique de X, de dimension $< n$.

4) Pour tout $x \in X \setminus B$ il existe un voisinage ouvert $U \subset X \setminus B$ tel que la restriction $f|U$ soit un isomorphisme sur $f(U) \subset Y \setminus D$.

Une conséquence de l'énoncé est que la restriction $f| : X \setminus B \longrightarrow Y \setminus D$ est un revêtement non ramifié. Pour le voir, il faut démontrer que cette restriction est propre. Pour cela, on peut utiliser l'argument de la fin de la démonstration de la proposition 5.1 ci-dessous. Le point est que, si $\varphi : X \longrightarrow Y$ est propre et si $K \subset Y$ est compact, alors $\varphi| : X \setminus \varphi^{-1}(K) \longrightarrow Y \setminus K$ est propre.

Note. Les ouverts $X \setminus B$ et $Y \setminus D$ sont denses et connexes. Voir le §7 sur les espaces normaux.

Définition 2.2 *L'entier d s'appelle le* **degré** *du morphisme fini f.*

Vocabulaire. Les sous-ensembles B et D ont divers noms dans la littérature. On dit parfois que $D \subset Y$ est le **discriminant** de f. On peut aussi dire que B est la **ramification en haut** tandis que D est la **ramification en bas**. On verra au §3 que f est un revêtement ramifié au sens des topologues. Mais je garderai la terminologie "morphisme (analytique) fini" pour la géométrie analytique et "revêtement ramifié" pour la topologie.

3. Revêtements ramifiés en topologie selon Karl Stein

La situation de départ est la suivante. Elle est volontairement un peu vague car nous n'arriverons à une définition satisfaisante qu'après quelques ajustements.

On a un espace topologique E, un sous-espace fermé $R \subset E$ et un revêtement fini (non ramifié) $g : Q \longrightarrow E \setminus R$ au-dessus de $E \setminus R$. On désire prolonger la projection g du revêtement en une application $f : \widehat{Q} \longrightarrow E$ sur E tout entier et l'on désire que ce prolongement soit unique. Si l'on ne prend pas certaines précautions, on ne peut pas espérer avoir unicité.

Voici un exemple simple, qui explique assez bien les conditions qui vont suivre. Soit Δ^2 le disque ouvert de rayon 1 dans $\mathbf{C}$, et soit $\check{\Delta}^2$ le disque épointé, c'est-à-dire privé de son centre. Soit $k \geq 2$ un entier et soient $\breve{\Delta}^2_i$ des disques épointés, pour $i = 1, \ldots, k$. Pour chaque entier i choisissons un entier $d_i \geq 1$. Soit $g_i : \breve{\Delta}^2_i \longrightarrow \breve{\Delta}^2$ un revêtement cyclique de degré d_i. Soit $\cup_i g_i : \cup_i \breve{\Delta}^2_i \longrightarrow \breve{\Delta}^2$ le revêtement d'espace total l'union disjointe des disques épointés et qui est égal à g_i sur $\breve{\Delta}^2_i$. Chaque g_i se complète de façon canonique en un revêtement ramifié du disque Δ^2_i sur Δ^2 en envoyant centre sur centre. Mais dans la réunion des Δ^2_i il y a beaucoup de façons de recoller les centres entre eux. On obtient ainsi des prolongements de $\cup_i g_i$ qui ne sont isomorphes en aucun sens. Bien sûr, dans l'exemple ci-dessus, la bonne façon de faire est de ne pas recoller les centres entre eux. Mais dans un cadre plus général, on ne peut pas dire les choses aussi simplement que cela !

Vocabulaire provisoire. L'application f est LE revêtement ramifié déterminé par le revêtement non ramifié g. Il est intuitivement plausible (c'est à cela que sert l'unicité) que f est déterminé par $R \subset E$ et par g à isomorphisme près.

La définition suivante est ce qui convient pour cet exposé. Dans [5] et surtout dans [9] on trouve une étude détaillée et bien plus générale du concept "ne pas séparer quelque part".

Définition 3.1 *Soit E un espace topologique et soit $A \subset E$ un fermé. On dit que A* **ne sépare E nulle part** *si, pour tout ouvert connexe non vide $U \subset E$ la différence $U \setminus (U \cap A)$ est non vide et (encore) connexe.*

Commentaire. Si B est une variété topologique de dimension m et si A est de dimension $\leq (m-2)$ alors A ne sépare B nulle part, essentiellement par dualité d'Alexander. En revanche revenons à l'exemple de la réunion $\cup_i \Delta^2_i$ des disques (pour $i = 1, \ldots, k$ avec $k \geq 2$) dont on identifie tous les centres en un point A pour obtenir un espace B. Alors ce point A sépare l'espace quotient B.

Venons-en à la définition de Stein d'un revêtement ramifié dans le cas topologique. Il existe des définitions plus générales, par exemple dans le cadre des spreads (voir le §5.2 ci-dessous) ; mais celle de Stein suffit à nos besoins.

Définition 3.2 *Soient X et Y des espaces topologiques connexes, localement connexes et localement compacts. Soit $f : X \longrightarrow Y$ une application propre. On dit que f* **est un revêtement ramifié à d feuilles** *s'il existe un sous-ensemble fermé $D \subset Y$, ne séparant Y nulle part, et tel que :*

(a) $f^{-1}(D) = B$ ne sépare X nulle part ;

(b) $f| \; X \setminus B \longrightarrow Y \setminus D$ est un homéomorphisme local ;

(c) chaque point $y \in Y \setminus D$ a une image réciproque $f^{-1}(y)$ qui se compose d'exactement d points distincts (nécessairement dans $X \setminus B$) ;

(d) l'image réciproque $f^{-1}(y)$ de n'importe quel point $y \in D$ est finie.

Il découle des définitions que f est une application ouverte et que n'importe quelle fibre $f^{-1}(y)$ a au plus d points. Pour les espaces topologiques qui satisfont les conditions de la définition, un homéomorphisme local propre est un revêtement non ramifié au sens usuel de la définition des revêtements. Par conséquent l'application décrite dans (b) est un revêtement non ramifié au sens usuel. Précisons qu'un point $y \in Y$ est un point de ramification (en bas) si y ne possède pas de voisinage au-dessus duquel f est un revêtement non ramifié. On dit aussi que f est ramifiée au-dessus de $y \in Y$. Par définition, $D \subset Y$ est l'ensemble des points de ramification en bas. Il se peut que D soit vide. Dans ce cas, bien évidemment, f est un revêtement non ramifié.

En ce qui concerne la topologie, le résultat de Stein est un théorème de prolongement des revêtements non ramifiés. C'est un théorème d'existence et aussi d'unicité. Voici l'essentiel du Satz 1 de Stein [12].

Théorème 3.1 *Données.*
(i) Un espace Y connexe, localement connexe et localement compact.
(ii) Un sous-espace fermé $D \subset Y$, ne séparant Y nulle part.
(iii) Un revêtement non ramifié $g : Z \longrightarrow Y \setminus D$ à d feuilles.
Alors il existe :
(1) Un espace X connexe, localement connexe et localement compact, qui contient Z comme un ouvert ;
(2) Une application $f : X \longrightarrow Y$ qui est un revêterment ramifié à d feuilles et telle que la restriction $f|\ Z = g$.
De plus f est ramifié au plus sur D.
Unicité :
Si $f' : X' \longrightarrow Y$ est une application jouissant des mêmes propriétés que f ci-dessus, il existe un homéomorphisme $\Phi : X \longrightarrow X'$ qui est l'identité sur Z et tel que $f' \circ \Phi = f$.

Indications sur la preuve de Stein

Il vaut la peine d'indiquer comment Stein s'y prend pour compléter le revêtement non ramifié. Le principe de la méthode peut s'adapter à des situations plus générales. Il utilise la notion de filtre sur un ensemble, notion due, sauf erreur, à Bourbaki. En voici l'essentiel, en suivant les notes d'un cours que Georges de Rham avait donné à Genève en 1956-1957.

Soit donc E un ensemble. Un **filtre** $\mathcal{F}$ sur E est la donnée d'un ensemble de parties de E qui satisfait les propriétés suivantes :
I. Toute partie de E qui contient une partie de E qui est un élément de $\mathcal{F}$ appartient à $\mathcal{F}$.
II. L'intersection de deux éléments de $\mathcal{F}$ appartient à $\mathcal{F}$.
III. L'ensemble vide n'appartient pas à $\mathcal{F}$ et E appartient à $\mathcal{F}$.

Commentaires.
(i) Il ne faut pas oublier que les éléments de $\mathcal{F}$ sont des parties de E. On peut donc, par exemple, prendre l'intersection de deux éléments de $\mathcal{F}$.
(ii) Ce qu'est **une base de filtre** est évident : un ensemble de parties $\mathcal{B}$ de E est une base du filtre $\mathcal{F}$ si $\mathcal{B} \subset \mathcal{F}$ et si tout ensemble de $\mathcal{F}$ contient un ensemble de $\mathcal{B}$.
(iii) Un ensemble $\mathcal{B}$ de parties de E est une base de filtre si les deux conditions suivantes sont satisfaites :

(a) l'intersection de deux éléments de $\mathcal{B}$ est un élément de $\mathcal{B}$.

(b) $\mathcal{B}$ n'est pas vide et l'ensemble vide n'appartient pas à $\mathcal{B}$.

Si E est un espace topologique et si $A \subset E$ est une partie de E, l'ensemble des voisinages de A est un filtre sur E.

Revenons à la donnée du Satz 1 de Stein. Nous allons indiquer comment Stein complète Z pour obtenir X.

Soit $y \in D$. Soit $U(y)$ un voisinage ouvert et connexe de y dans Y. Soit $V(y) = U(y) \setminus U(y) \cap D$. Comme D ne sépare Y nulle part, $V(y)$ est ouvert, non vide et connexe. On considère alors l'image réciproque $g^{-1}(V(y)) \subset Z$. C'est un ouvert de Z et la restriction de g à cet ouvert est un homéomorphisme local sur $V(y)$. Une composante connexe de $g^{-1}(V(y))$ est appelée une **composante distinguée** de Z et notée $\widetilde{V}(y)$.

Pour $y \in D$ toujours fixé, les voisinages tels que $U(y)$ constituent une base du filtre des voisinages de y. Les $V(y)$ forment une base de filtre $\mathcal{F}_y$ sur $Y \setminus D$. Nous considérons alors les filtres $\widetilde{\mathcal{F}_y}$ sur Z qui ont une base formée des composantes distinguées $\widetilde{V}(y)$. Comme g est un revêtement à d feuilles, il est clair que pour $y \in D$ fixé, on a au plus d tels filtres. On note ces filtres $\widetilde{y}_1, \ldots, \widetilde{y}_r$ avec $1 \leq r \leq d$. Par définition $\widetilde{y}_1, \ldots, \widetilde{y}_r$ sont les points de X qui se trouvent au-dessus de y. L'ensemble des points que l'on ajoute ainsi à Z pour y parcourant D est noté B. Par définition $X = Z \cup B$. On étend g en $f : X \longrightarrow Y$ en posant $f(\widetilde{y}_j) = y$ pour $1 \leq j \leq r$.

La suite de la construction consiste à munir X d'une topologie qui satisfait toutes les conditions requises. Cette topologie est obtenue ainsi. Soit $\widetilde{V}(y)$ une composante distinguée. Soit S l'ensemble des points de B qui ont ce $\widetilde{V}(y)$ parmi les éléments de leur base de filtre. On note $\widetilde{W}(y)$ la réunion $\widetilde{V}(y) \cup S$. Alors par définition les ouverts de X sont les ouverts de Z et les $\widetilde{W}(y)$. La vérification que cette topologie a les propriétés requises ne présente pas de difficulté majeure.

La démonstration de l'unicité demande un argument ad hoc, exprimé dans le lemme (Hilfsatz) 2 de [12]. Le fait que la ramification ne sépare nulle part est essentiel.

4. Prolongement de revêtements dans le cas analytique

Je vais suivre principalement l'article de Cartan [4] et celui de Grauert-Remmert [6]. Le point de départ de ces derniers auteurs est le Satz 1 de Stein, adapté à la situation qui les concerne. Il est énoncé dans leur Satz 8, dont voici le contenu.

Théorème 4.1 *Données.*
(i) Une variété analytique (lisse donc) Y connexe.
(ii) Un sous-ensemble analytique $D \subset Y$ de codimension ≥ 1.
(iii) Un revêtement topologique non ramifié à d feuilles $g : Z \to Y \setminus D$.
Alors il existe :
(1) Un espace topologique X connexe, localement connexe et localement compact, contenant un fermé B ne séparant X nulle part et tel que $X \setminus B = Z$.
(2) Une application $f : X \longrightarrow Y$ qui est un revêtement ramifié (au sens topologique bien sûr) à d feuilles prolongeant g.

Commentaires.
1. Il y a unicité, comme dans le Satz 1 de Stein.
2. L'ouvert $Y \setminus D$ est une variété analytique ; comme $g : Z \longrightarrow Y \setminus D$ est un revêtement non ramifié, il est facile de munir Z d'une structure de variété analytique de façon que g soit un isomorphisme local. C'est ce que l'on fait.
3. Ce que l'on obtient alors est appelé par Grauert et Remmert une "analytische Überlagerung". Attention : ce qui est analytique consiste en la donnée en bas ($D \subset Y$) et en le revêtement non ramifié $g : Z \longrightarrow Y \setminus D$. Mais pour l'instant X n'est pas un espace analytique. Le but de l'article [6] est précisément de munir X d'une structure d'espace analytique normal, qui prolonge la structure lisse sur Z.

Indications sur la preuve du théorème 1.1 (existence)

En gros, la situation est la suivante :

X est un espace topologique localement compact ; $B \subset X$ est un fermé qui ne sépare X nulle part et l'ouvert $X \setminus B$ est une variété analytique de dimension complexe n. Il s'agit d'étendre la structure sur Z en une structure normale sur X, de façon à ce que B soit un sous-espace analytique de dimension $< n$.

Pour atteindre ce but, on munit X d'une structure d'espace annelé, en construisant un faisceau $\mathcal{S}_X$ sur X. Ensuite, la démonstration consiste à prouver que $\mathcal{S}_X$ est une structure analytique ayant les propriétés voulues. Il faut admettre que la démonstration de [6] est difficile. Je vais me contenter d'en donner l'idée générale, en suivant Cartan, qui a publié deux articles sur la question. Voir [4]. Celui du Séminaire reprend celui publié aux Mathematische Annalen, avec quelques variations. Voir par exemple le Théorème 2 p.11-06 du Séminaire. La présentation de Cartan est très claire.

Pour définir $\mathcal{S}_X$ en un point $z \in Z$ on n'a pas le choix, puisque Z est déjà muni d'une structure analytique (lisse).

Soit donc $b \in B$. Soit V un voisinage ouvert (connexe) de b dans X. On définit le faisceau $\mathcal{S}_X$ sur V comme étant formé des fonctions continues $\varphi : V \longrightarrow \mathbf{C}$ qui sont holomorphes en tout point $z \in V \cap Z$. Cartan observe que, si l'on désire que X soit normal, on est obligé de procéder ainsi, puisque les espaces normaux jouissent précisément de cette propriété (si l'on sait que B est un sous-ensemble analytique, ou plus généralement "thin"). Voir l'Appendice 2 ci-dessous. Je regrette de ne pas pouvoir en dire plus sur l'existence.

Remarquons que, une fois que l'on sait que la structure analytique sur X est normale et que $B \subset X$ est un sous-espace analytique, la projection $f : X \longrightarrow Y$ du Satz 8 est automatiquement analytique. L'argument est le même que dans les lignes qui suivent ci-dessous.

Indications sur la preuve du théorème 1.2 (unicité)

La preuve est une conséquence immédiate du fait que les structures sont normales. En effet, considérons l'homéomorphisme $\Phi : X \longrightarrow X'$ du théorème 3.1. La restriction $\Phi| : Z \longrightarrow Z$ est l'identité. En particulier, Φ est continue et sa restriction aux points lisses est holomorphe. Par conséquent, Φ est analytique. De même l'inverse $\Phi' : X' \longrightarrow X$ est analytique pour les mêmes raisons. Et donc Φ est un isomorphisme analytique.

Il est intéressant de lire ce que dit Remmert dans [11] au §4.1. Un "analytically ramified covering" est une "analytische Überlagerung" dont nous venons de parler.

Remarques. (1) Les prolongements topologiques à la Stein se comportent bien par changement de base, en prenant certaines précautions. C'est ce qu'affirme la proposition suivante.

Proposition 4.1 *Soient :*

(a) $D \subset Y$ *;* $g : Z \longrightarrow Y \setminus D$ *et* $f : (X, B) \longrightarrow (Y, D)$ *comme dans le Satz 1.*

(b) $\psi : (Y^+, D^+) \longrightarrow (Y, D)$ *un homéomorphisme.*

Alors :

(c) le pull-back $g^+ : Z^+ \longrightarrow Y^+ \setminus D^+$ *de* $g : Z \longrightarrow Y \setminus D$ *se prolonge, de façon essentiellement unique, en un revêtement ramifié* $f^+ : (X^+, B^+) \longrightarrow (Y^+, D^+)$

(d) l'homéomorphisme ψ *se relève en un homéomorphisme* $\Psi : (X^+, B^+) \longrightarrow (X, B)$ *tel que l'on ait l'égalité* $f \circ \Psi = \psi \circ f^+$.

Mutatis mutantis, on a un résultat analogue pour le théorème de prolongement de Grauert-Remmert dans le cadre analytique. La condition importante est que $\psi : (Y^+, D^+) \longrightarrow (Y, D)$ doit être un isomorphisme analytique.

(2) Tout ce passe bien du côté des groupes de Galois. On a en effet la proposition suivante.

Proposition 4.2 *Soient $D \subset Y$ et $g : Z \longrightarrow Y \setminus D$ comme dans le Satz 1. Supposons que le revêtement (non ramifié) g soit Galoisien de groupe G. Alors le revêtement ramifié $f : (X, B) \longrightarrow (Y, D)$ est Galoisien de groupe G.*

On a un énoncé analogue dans le cadre analytique. En ce qui concerne les surfaces de Riemann, ceci est un procédé élégant et ancien pour construire des surfaces avec automorphismes. Voir, par exemple, le problème de Galois inverse. Un point de départ est que, étant donné un groupe fini G, le théorème d'existence de Riemann permet de construire un revêtement galoisien $\Sigma^2 \longrightarrow S^2$ ayant G pour groupe de Galois. Pour cela, on construit tout d'abord un revêtement galoisien non ramifié de S^2 privée d'un nombre adéquat de points. Ensuite on le complète. Hélas, le théorème d'irréductibilité de Hilbert ne permet pas de conclure qu'il existe une extension finie des rationnels ayant G pour groupe de Galois.

(3) J'achève ce §4 en faisant quelques commentaires sur divers avatars des énoncés précédents.

(i) Dans la situation de la Proposition 4.1 les espaces X et X^+ sont homéomorphes mais les structures analytiques ne sont en principe pas isomorphes. Un changement de base topologique nous conduit donc à une situation de modules d'espaces analytiques. Dans la version où le changement de base ψ est analytique, alors X et X^+ sont analytiquement isomorphes. Dans ce contexte, il y a une situation (certes rare, mais remarquable) où l'existence d'un changement de base topologique entraine l'existence d'un ψ analytique. C'est ce qu'il se passe pour les dessins d'enfants, dont le point de départ est la donnée suivante. L'espace Y est la sphère de Riemann S^2 et $D \subset S^2$ est un ensemble de 3 points. Or, le groupe des automorphismes analytiques de S^2 agit transitivement sur les sous-ensembles de 3 points. Par conséquent, la donnée de ce qui est nécessaire pour déterminer topologiquement un revêtement ramifié sur 3 points détermine la structure analytique de la surface de Riemann obtenue comme espace total du revêtement. Cette donnée topologique est un dessin d'enfants. Il est vraiment surprenant qu'une donnée aussi "molle" détermine une structure aussi "délicate". La suite de cette superbe histoire est racontée, par exemple, dans le séminaire Bourbaki de Joseph Oesterlé No 907 (2001-2002) et dans plusieurs textes de Leila Schneps. Il est extraordinaire que le groupe de Galois de $\bar{\mathbf{Q}}/\mathbf{Q}$ agisse (Alexandre Grothendieck !) sur l'ensemble des dessins d'enfants.

(ii) Un autre exemple d'utilisation de ces résultats d'existence et d'unicité de prolongements se trouve dans le livre de Gottfried Barthel, Friedrich Hirzebruch et Thomas Höfer [1]. L'objectif est de construire des surfaces complexes (variétés analytiques complexes, donc lisses, compactes, de dimension complexe 2) dont les nombres de Chern ont certaines propriétés. Sans entrer dans les détails, la stratégie est la suivante. On choisit pour Y l'espace projectif $P^2(\mathbf{C})$. Le sous-espace $D \subset P^2(\mathbf{C})$

est donc une courbe projective. Pour maitriser les calculs, on choisit pour D un arrangement de droites. On se donne ce qu'il faut pour construire un revètement fini du complémentaire de D. On le complète par un espace normal. Ce dernier n'a alors que des singulariés isolées, résolues finalement par la méthode de Hirzebruch-Jung. Il faut de la virtuosité pour mener à bien ce programme.

5. Un peu d'histoire

5.1. L'uniformisation des fonctions algébriques

L'article de Reinhold Remmert [11] est une référence très documentée sur l'histoire de l'uniformisation des fonctions algébriques. Traditionnellement on part d'un polynôme

$$P(z,w) = w^d + a_1(z)w^{d-1} + \cdots + a_{d-1}(z)w + a_d(z)$$

où les $a_i(z)$ sont des polynômes à n variables $z = (z_1, \ldots, z_n)$ à coefficients complexes. Si on fixe le point $z \in \mathbf{C}^n$, quel que soit le nombre n de variables l'équation polynomiale $P(z,w) = 0$ possède **en général** d racines distinctes. Tout le sel de l'affaire est dans le "en général". De toutes façons on considère que w est une fonction de z ayant en général d valeurs. On dit qu'elle est multiforme de degré d. Pendant une bonne partie du 19ème siècle on ne considéra que le cas $n = 1$. Typiquement la fonction $\sqrt[2]{q(z)}$ où $q(z)$ est un polynôme de degré 3 à une variable est multiforme. C'est une telle fonction qui intervient dans le calcul de la longueur de l'arc de l'ellipse. Cela pose un problème : comment intégrer une fonction multiforme ? (En fait, on devrait parler de forme holomorphe ou méromorphe multiforme.) La réponse est (en langage d'aujurd'hui) : la fonction devient (ou "est") uniforme sur un revêtement ramifié de la sphère de Riemann S^2. La ramification vient précisément des valeurs de z pour lesquelles le polynôme possède moins de d racines. On ne peut pas éviter l'existence de ramification. Les prédécesseurs de Riemann savaient en fait beaucoup de choses sur ce qu'il se passe autour d'un point de ramification (pour $n = 1$!). Voir par exemple l'article de Puiseux, publié en 1850 [10]. D'ailleurs, je trouve que l'Histoire est injuste à son sujet.

Faisons un saut dans le temps en utilisant un langage plus récent, qui est précisé dans l'alinéa juste ci-dessous. Les points de ramification ne causent pas vraiment de difficulté pour les fonctions algébriques **à une variable** pour la raison suivante. L'ensemble des points de ramification en bas est formé de points isolés. Soit Q un tel point. Soit Δ un petit disque ouvert centré en Q et ne contenant que Q comme point de ramification. Soit $\check{\Delta}$ le disque épointé. Alors $\Pi_P^{-1}(\check{\Delta}) \longrightarrow \check{\Delta}$ est un revêtement non ramifié à d feuilles, d'espace total non nécessairement connexe. Comme le groupe fondamental du disque épointé est isomorphe aux entiers rationnels, la monodromie de ce revêtement est déterminée par une permutation σ de l'ensemble $\{1, \ldots, d\}$. Décomposons σ en produits de cycles deux à deux disjoints : $\sigma = c_1, \ldots, c_i, \ldots, c_k$

pour un certain entier $k \leq d$. L'ensemble des cycles correspond bijectivement à l'ensemble des composantes connexes de $\Pi_P^{-1}(\breve{\Delta})$. Désignons ces composantes connexes par $\breve{\Delta}_1, \ldots, \breve{\Delta}_i, \ldots, \breve{\Delta}_k$. Le revêtement $\breve{\Delta}_i \longrightarrow \breve{\Delta}$ est cyclique d'ordre d_i. Il est isomorphe (via un isomorphisme φ_i) au revêtement $\breve{D}_i^2 \longrightarrow \breve{D}^2$ défini par $z \mapsto z^{d_i}$ (où D^2 désigne le disque ouvert de rayon 1 dans $\mathbf{C}$). Considérons le quotient de l'union disjointe de $\cup_i D_i^2$ avec $\Pi_P^{-1}(\breve{\Delta})$ obtenu en recollant $\cup_i D_i^2$ à $\Pi_P^{-1}(\breve{\Delta})$ à l'aide des isomorphismes φ_i. Ce quotient est la surface de Riemann lisse cherchée. La raison essentiellle pour laquelle l'opération est aussi facile est que la monodromie autour de la ramification est très simple.

A partir de la fin du 19ème siècle on s'est posé la question : que se passe-t-il si $n \geq 2$? Par exemple si $n = 2$? Bien sûr, le lieu naturel qui uniformise la fonction algébrique est l'hypersurface Σ_P de $\mathbf{C}^{n+1}$ d'équation $P(z, w) = 0$. Dans le langage que nous utilisons ici la projection naturelle $\Pi_P : \Sigma_P \longrightarrow \mathbf{C}^{n+1}$ est un morphisme analytique fini, dont le lieu de ramification en bas est le lieu des zéros du discriminant du polynôme $P(z; w)$, considéré comme polynôme en la variable w. Mais vers 1900 ces concepts n'étaient pas tous disponibles. A cette époque, on parlait d'espace (ou de domaine) de Riemann pour qualifier l'application $\Pi_P : \Sigma_P \longrightarrow \mathbf{C}^n$. Wilhelm Wirtinger (élève de Felix Klein) fut un des premiers à réaliser que localement il y a une difficulté déjà pour $n = 2$. Wirtinger et certains prédécesseurs (par exemple Poul Heegaard dans sa thèse de 1898) observent que l'ensemble de ramification en bas est une courbe. Plaçons-nous en un point singulier a de la courbe discriminant. Pour simplifier la discussion, supposons que Σ_P est normale. Soit b un point au-dessus de a. Alors, Σ_P est localement en b le cône de base l'espace total d'un revêtement de la sphère S^3 ramifié sur un entrelacs. Cet entrelacs est l'intersection du bord d'un petit voisinage sphérique de a avec la courbe discriminant. Il n'y a en principe aucune raison pour qu'un tel point soit uniformisable. (Pendant très longtemps, on a appelé **uniformisable** un point de l'hypersurface qui est lisse.) Wirtinger a souvent posé la question de savoir quels sont les entrelacs qui peuvent être de tels ensembles de ramification et quels sont les revêtements finis qui leur correspondent. La détermination de ces entrelacs (dits algébriques) fut effectuée par son élève Karl Brauner et par Werner Burau.

C'est ce genre de questions qui est le point de départ de l'article de Heinrich Behnke et Karl Stein [2], dont on peut considérer qu'il a eu une longue descendance. Tout d'abord, ils observent que, si l'on peut compactifier $\mathbf{C}^2$ topologiquement ou différentiablement en la sphère S^4 on ne peut pas le faire analytiquement. La situation est donc plus compliquée que dans le cas $n = 1$, où l'on peut compactifier analytiquement $\mathbf{C}$ en ajoutant un seul point à l'infini (ce que l'on fait habituellement). On se résoud donc à laisser l'hypersurface Σ_P telle qu'elle est. Mais il y a plus sérieux. Dans l'hypersurface Σ_P il peut se trouver des points vraiment non uniformisables. L'exemple probablement le plus simple est celui que donnent Behnke-Stein : Il s'agit de l'origine de $\mathbf{C}^3$ pour l'hypersurface d'équation $P(z, w) = w^2 - z_1 z_2 = 0$. Elle est normale et, localement, Σ_P est le cône sur l'espace projectif réel $P^3(\mathbf{R})$. Cet exemple avait déjà été découvert par Poul Heegaard dans sa thèse. Il fut d'ailleurs à l'origine d'une controverse avec Henri Poincaré, sur l'énoncé de la dualité de Poincaré lorsque l'homologie a de la torsion. Manifestement, on ne peut pas, par une manipulation ponctuelle, transformer Σ_P en une variété, comme ce fut le cas par normalisation lorsque $n = 1$. Friedrich Hirzebruch a proposé dans sa thèse soutenue à Münster (!) en 1950 de résoudre les singularités lorsque $n = 2$, en suivant la méthode, dite aujourd'hui de Hirzebruch-Jung, qui fait précisément l'objet de sa thèse. On obtient ainsi une variété $\widetilde{\Sigma}_P$ et un morphisme propre $\widetilde{\Pi}_P : \widetilde{\Sigma}_P \longrightarrow \Sigma_P$ (mais qui n'est pas à fibres finies) tel que la composition $\Pi_P \circ \widetilde{\Pi}_P$ uniformise la fonction algébrique. Voir aussi le séminaire Bourbaki 84 en décembre 1953 de Cartan.

Plus généralement, il s'agit de comprendre le morphisme Π_P pour n quelconque. Certainement, il s'agit d'une application continue, propre, entre espaces localement compacts, à fibres finies. Elle ressemble à un revêtement topologique traditionnel sans en être vraiment un. De plus, une partie des données est analytique (et même algébrique). Dans leur article de 1951, Behnke-Stein proposent de trianguler la situation. Que cela soit possible avait déjà été démontré par A.B. Brown et B.O. Koopman en 1933. Une fois triangulée, Σ_P est une pseudo-variété. C'est, en topologie combinatoire, ce qui peut se faire de mieux si l'on n'a pas une variété. Salomon Lefschetz avait aussi vanté les mérites des pseudo-variétés (appelées circuits) pour décrire la topologie des variétés algébriques projectives complexes (comme il s'agit de variétés algébriques, elles peuvent être singulières). Le §5.4 ci-dessous est consacré aux pseudo-variétés (triangulées). L'analogie avec certains aspects des espaces analytiques est frappante. Du coup, la projection $\Pi_P : \Sigma_P \longrightarrow \mathbf{C}^n$ est simpliciale. Dans ce cadre-là, il n'est pas très difficile de définir ce qu'est un revêtement ramifié. Il faut néanmoins observer que dans cette situation la ramification en haut peut séparer localement. Les revêtements ramifiés de Behnke-Stein sont donc plus généraux que ceux décrits par Stein. Bien sûr, on peut (on doit) considérer que les triangulations sont comme une sorte de corps étranger dans le sujet et qu'il vaudrait mieux utiliser des méthodes plus intrinsèques. C'est exactement à cette préoccupation que répond l'article de Stein dont nous avons parlé dans le §3.

5.2. Les spreads de Ralph Fox et José Montesinos

C'est à peu près à la même époque (milieu des années 1950-1960) que Ralph Fox a proposé une définition purement topologique des revêtements ramifiés. Comme pour Stein, son ambition est de se passer des triangulations, mais pour de toutes autres raisons. En ce temps-là, beaucoup des invariants construits en topologie dépendaient du choix d'une triangulation et n'étaient donc pas de véritable invariants topologiques. En particulier c'était le cas en théorie des noeuds, qui faisait un grand usage des revêtements de la sphère S^3 ramifiés sur un noeud on un entrelacs. Or ces revêtements étaient définis à l'aide d'une triangulation et il s'agissait donc d'éliminer leur usage. Il semble qu'il n'y a pas eu de contacts entre la géométrie analytique (Paris et Münster) d'une part et la topologie purement topologique (Princeton en l'occurrence) d'autre part et que les deux théorie se soient développées indépendamment.

Fox a introduit la notion de "spread" (que l'on pourrait traduire par "étendue", mais je vais conserver le terme spread), qui généralise grandement les revêtements ramifiés. Voir [5]. Plus tard, José Montesinos a magistralement placé les spreads dans un contexte encore plus général. Voir [9]. Voici, en gros, de quoi il s'agit.

Les espaces que l'on considère sont supposés seulement T_1, ce qui signifie que chaque point est un fermé dans la terminologie introduite par Felix Hausdorff. C'est nettement plus faible que T_2 qui signifie séparé. On ne suppose pas au départ que les espaces sont à base dénombrable, ou métrisables. Nous sommes donc très loin des espaces localement compacts utilisés dans le §2.

Définition 5.1 *Soient X et Y deux espaces T_1. Une application continue $f : X \longrightarrow Y$ est un* **spread** *si la condition suivante ★ est satisfaite :*

★ *une base de la topologie de X est obtenue en considérant les composantes connexes des images réciproques par f des ouverts de Y.*

Concrètement cela signifie que pour tout $x \in X$ et tout ouvert $U \ni x$, il existe un ouvert $W \subset Y$ et une composante connexe $\widehat{W}$ de $f^{-1}(W)$ telle que $x \in \widehat{W} \subset U$.

Commentaire. Tout revêtement traditionnel satisfait la condition ★. Il était audacieux de la part de Fox de prendre comme point de départ cette unique propriété. Fox et plus tard Montesinos jouent le jeu à fond et n'imposent que peu ou pas de condition supplémentaire aux espaces topologiques. Par conséquent, l'étude des spreads est vite très technique et le néophyte (j'en suis un !) est un peu perdu.

Quelques conséquences faciles. Comme Y est T_1 les fibres $f^{-1}(y)$ pour $y \in Y$ sont fermées. La condition ★ implique facilement que X est localement connexe. La condition ★ implique aussi qur les fibres sont totalement discontinues : leurs composantes connexes sont réduites à un point. Mais il y a des exemples où une fibre est un Cantor. Dans la littérature, une application telle que les fibres sont totalement discontinues est appelée **light**.

Noter que l'on n'impose pas que l'application f soit surjective. En général un spread n'est ni une application ouverte ni une application fermée. Ceci indique que les spreads sont loin des revêtements. L'article cité de José Montesinos donne beaucoup d'exemples intéressants, qui montrent que le monde des spreads est plein de surprises. On jugera du degré de généralité de la notion de spread en consultant le §6 de l'article de Fox, consacré aux spreads dans le cadre des complexes simpliciaux. Une application simpliciale $f : K_1 \longrightarrow K_2$ est un spread si et seulement si, pour tout simplexe σ de K_1, l'image $f(\sigma)$ est un simplexe de K_2 de même dimension que σ. Autrement dit, f est non dégénérée au sens des applications simpliciales. C'est tout.

Voici quelques résultats que je trouve rassurants, mais tout de même loin des revêtements.

1) Supposons que $f : X \longrightarrow Y$ est light, avec X un espace séparé et Y un espace T_1. Si f est fermée et à fibres compactes (ceci est souvent pris comme définition d'une application propre, lorsque les espaces ne sont pas localement compacts) alors f est un spread.

2) Si X et Y sont des variétés séparées, une application $f : X \longrightarrow Y$ est un spread si et seulement si elle est light.

Fox et Montesinos ont besoin d'une notion qui est proche de “ne sépare nulle part”.

Définition 5.2 *Soit E un espace topologique et soit $W \subset E$ un ouvert (Fox et Montesinos ne prennent même pas cette restriction). On dit que W* **est localement connexe dans** *E si :*

(1) W est dense dans E.

(2) il existe une base d'ouverts $\{U_j\}_{j \in J}$ de E telle que pour tout U_j l'intersection $U_j \cap W$ est connexe.

Comparer avec la définition de “ne sépare nulle part” en prenant $W = E \setminus A$.

Les deux auteurs ont besoin de cette notion pour faire entrer les revêtements ramifiés dans le cadre des spreads. Mais de toutes façons leur définition est bien plus générale que celle que j'ai présentée dans le §3, due à Stein. Je renvoie les lecteurs à l'article très documenté de Montesinos, qui contient une bibliographie détaillée sur les spreads.

5.3. Les espaces analytiques complexes

Au début des années 1950, il était donc clair que les variétés analytiques ne sont pas suffisantes pour résoudre certains problèmes qui se présentent en géométrie complexe. D'autres objets sont nécessaires. Dans son article [11] Remmert fait une observation qui m'a surpris (ce qui témoigne de mon ignorance) : vers 1950, même les variétés analytiques (lisses !) étaient peu étudiées. Ceci est confirmé par Cartan dans [3].

Le concept qui a tout déclenché est la notion de faisceau, due à Jean Leray et magistralement étudiée et exploitée par Henri Cartan et Jean-Pierre Serre. Tant qu'il s'agit de variétés (différentiables, analytiques, affines, hyperboliques, ...) on peut définir la structure désirée par des cartes, avec la condition que les changements de cartes doivent appartenir à un pseudo-groupe donné. Un livre publié vers 1930, dû à Oswald Veblen et Henry Whitehead, a bien expliqué comment procéder. Malheureusement, ce livre a paru trop compliqué et n'a pas eu une grande influence, me semble-t-il. Mais les changements de cartes ne fournissent pas la recette adéquate pour définir les espaces analytiques. C'est là que les faisceaux interviennent, pour recoller les structures locales. Dans les deux appendices, j'esquisse grossièrement comment cela fonctionne.

5.4. Les pseudo-variétés triangulées

On considère des complexes simpliciaux K, pas nécessairement finis, mais localement finis. On les munit de la topologie faible : un sous-ensemble de K est, par définition, fermé si et seulement si son intersection avec chaque simplexe est fermée. La condition "localement fini" est équivalente à "localement compact". Avec la topologie faible, un complexe simplicial (on devrait dire l'espace topologique sous-jacent) est paracompact.

Définition 5.3 *Un complexe simplicial M est une* **pseudo-variété** *de dimension (topologique) m si les trois conditions suivantes sont satisfaites :*

(1) M est **purement de dimension** *m ; explicitement, tout simplexe de dimension $< m$ de M est face d'au moins un simplexe de dimension m.*

(2) Tout simplexe de dimension $(m-1)$ est face d'exactement deux simplexes de dimension m.

(3) Etant donnés deux simplexes de dimension m, on peut passer de l'un à l'autre par une suite finie de simplexes adjacents de dimension m et $(m-1)$ en alternance.

Commentaires. Notons M_0 l'ensemble des points de M qui ont un voisinage homéomorphe à un ouvert de $\mathbf{R}^m$. Par construction M_0 est un ouvert dense, constitué des points lisses. Le complémentaire $S(M)$ est le sous-complexe des **points singuliers de** M. La condition (2) dit que $S(M)$ est de codimension ≥ 2. La condition (3) dit que M_0 et M sont connexes. Autrement dit, M est irréductible. Si on le désire, on peut se passer de la condition 3). Ce qui est vraiment caractéristique des pseudo-variétés se trouve dans les conditions 1) et 2). Mais j'admettrai la condition d'irréductibilité.

Définition 5.4 *Soit M une pseudo-variété de dimension m. Soit $x \in M$. On dit que M est* **localement irréductible en** *x si x possède un système fondamental de voisinages ouverts $\{V_j\}$ pour $j \in J$ tels que $V_j \cap M_0$ soit connexe.*

Comme on n'est pas concernés par le prolongement de fonctions holomorphes (voir l'appendice 2), on dit souvent que M est normale, au lieu de localement irréductible. Si M est une pseudo-variété normale et si $K \subset M$ est un sous-complexe de codimension ≥ 2, alors K ne sépare M nulle part.

Behnke-Stein proposent une définition de revêtement ramifié d'une pseudo-variété (pas nécessairement normale, comme nous l'avons vu plus haut) sur un ouvert de $\mathbf{C}^n$. En général, les revêtements ramifiés entre pseudo-variétés normales fonctionnnent bien, comme l'indique la proposition ci-dessous. Il y a des articles passionnants de Clint McCrory et Dennis Sullivan sur les pseudo-variétés normales, en particulier sur **la normalisation des pseudo-variétés**.

En quelques mots, la normalisation d'une pseudo-variété M permet de construire une pseudo-variété normale $\widehat{M}$ et une application simpliciale $\nu : \widehat{M} \longrightarrow M$ qui jouit des propriétés suivantes :

i) La restriction $\nu|\ \nu^{-1}M_0 \longrightarrow M_0$ est un isomorphisme linéaire par morceaux. Ici, M_0 désigne l'ouvert dense des points de M où M est localement irréductible.

ii) L'application ν est propre à fibres finies.

En d'autres termes, ν est un revêtement à une feuille ! Mais comme M n'est pas normale, on peut avoir plus d'un point au-dessus d'un point de la ramification. L'exemple standard consiste à prendre pour $\widehat{M}$ la sphère S^2 et pour ν la projection sur la quotient de S^2 obtenu en identifiant deux points de S^2.

Il est intéressant de constater que les pseudo-variétés ont connu un regain de popularité grâce à l'homologie d'intersection de Marc Goresky et Robert MacPherson. Dans leur premier article, les pseudo-variétés sont triangulées. Par la suite, les triangulations ont été éliminées au profit des faisceaux, selon une loi naturelle de l'évolution en mathématiques.

C'est un théorème connu des spécialistes que les espaces analytiques peuvent se trianguler. On obtient alors une pseudo-variété de dimension (réelle) $2n$ si l'on triangule un espace analytique équidimensionel de dimension (complexe) n. De plus si l'espace analytique est normal au sens de la géométrie analytique, alors la pseudo-variété est normale.

Terminons cette partie PL par la proposition suivante, que l'on peut considérer comme une version allégée du théorème de Cartan du §2. Elle certainement bien connue des spécialistes des "light open mappings". Certes, elle ne concerne qu'une situation assez particulière, mais elle a le mérite de rendre visibles plusieurs notions que nous utilisons.

Proposition 5.1 *Soient M et N deux pseudo-variétés connexes, compactes,* **normales***, de même dimension m et triangulées. Soit $f : M \longrightarrow N$ une application simpliciale ouverte. Alors f est un revêtement ramifié au sens de Stein.*

L'énoncé est également vrai si M et N ne sont pas compactes, et si l'on ajoute alors l'hypothèse que f est propre. D'autre part, observons que la réciproque de l'énoncé est vraie, car nous avons vu au §3 que la projection d'un revêtement ramifié à la Stein est ouverte.

Preuve de la proposition

Commençons par observer que f est surjective, puisque son image est à la fois ouverte et fermée. Soit U l'intérieur d'un simplexe de dimension m de M. Alors $f(U)$ est l'intérieur d'un simplexe de même dimension de N, puisque f est ouverte et simpliciale. Par conséquent f est non dégénérée, puisque tout simplexe de M est face d'au moins un simplexe de dimension m. Par conséquent les fibres de f sont finies.

Introduisons la notation suivante. M_0 désigne le complémentaire dans M du squelette de dimension $(m-2)$. M_0 est connexe, d'après l'hypothèse d'irréductibilité des pseudo-variétés. C'est la réunion disjointe de l'intérieur des simplexes de M dimension m ou $(m-1)$. Dans ce qui suit, quand on parlera d'un simplexe de M_0 il s'agira d'un simplexe dans ce sens. Deux tels simplexes sont adjacents si l'intersection de leur adhérence est non vide. Mutanta mutantis pour N_0. Comme f conserve la dimension des simplexes, la restriction de f à M_0 induit une application surjective et ouverte $f_0 : M_0 \longrightarrow N_0$.

Claim. L'application $f_0 : M_0 \longrightarrow N_0$ est un revêtement non ramifié.

Il suffit de revenir à la définition de Stein ci-dessus au §3 pour voir que le Claim implique la proposition puisque le $(m-2)$-squelette d'une pseudo-variété normale de dimension m ne sépare nulle part. C'est ici que l'hypothèse de la normalité des pseudo-variétés intervient.

Preuve du Claim

Soit $P \in M_0$. Si P appartient à un simplexe μ de dimension m, la restriction f_0 de f à ce simplexe est localement un homéomorphisme puisque f est ouverte et simpliciale. Supposons donc que P appartient à un simplexe η de dimension $(m-1)$ de M_0. Soient μ_1 et μ_2 les deux simplexes de dimension m qui sont adjacents à η. Alors f_0 envoie ces deux simplexes sur des simplexes distincts de N_0, adjacents à $f_0(\eta)$. En effet, si ce n'était pas le cas ces deux simplexes auraient même image. L'application f serait donc ce que l'on appelle un pli (en anglais "a fold") le long de η. Mais un pli n'est pas une application ouverte.

Il reste à démontrer que f_0 est propre, car un homéomorphisme local et propre est un revêtement non ramifié. Ici, il faut prendre garde car la restriction d'une application propre à un ouvert n'est pas nécessairement propre. Dans notre cas, on peut raisonner ainsi. Soit M_0^* le compactifié d'Alexandroff de l'espace localement compact M_0. Un lemme facile sur les applications propres entre espaces localement compacts dit qu'une application continue est propre si et seulement si son extension ensembliste aux compactifiés (définie en envoyant ∞ sur ∞) est continue à l'infini. Or, on peut obtenir M_0^* en identifiant le $(m-2)$-squelette de M en un point, qui devient le point à l'infini de M_0^*. Comme f conserve les squelettes, elle passe aux quotients et induit une application continue $f_o^* : M_0^* \longrightarrow N_0^*$.

Ceci achève la preuve du claim et donc de la proposition.

Observations. 1) Il découle de la preuve que l'ensemble de ramification minimal en bas est contenu dans le $(m-2)$-squelette de N.

2) Attention. Si M et N sont des vraies variétés, il ne faut pas en conclure qu'une application simpliciale et ouverte est un revêtement non ramifié. La preuve précédente indique que la ramification en bas est contenue dans le $(m-2)$-squelette de N.

6. Appendice 1 : Espaces analytiques

Une superbe présentation de ce sujet est faite par Henri Cartan dans [3].

Encore une fois, je me restreins aux espaces analytiques sur le corps de complexes. Leur étude se fait en deux étapes.

Tout d'abord on se donne un domaine $U \subset \mathbf{C}^k$ et un point $a \in U$. On considère l'anneau des germes de fonctions holomorphes (= analytiques) en a. Si a est l'origine de $\mathbf{C}^k$ il s'agit de l'anneau $\mathbf{C}\{z_1, \ldots, z_k\}$ des séries convergentes en les variables $(z_1, \ldots, z_k)$. Cet anneau a des propriétés algébriques remarquables : il est local, noethérien et factoriel. On établit traditionnellement ces propriétés par le théorème de préparation de Weierstrass-Rückert.

De là on passe aux sous-ensembles analytiques dans le domaine $U \subset \mathbf{C}^k$. Par définition, il s'agit d'un sous-ensemble $A \subset U$ localement fermé tel que, pour tout point $a \in A$ il existe un voisinage ouvert V de a dans U de telle façon que $A \cap V$ soit égal au lieu des zéros d'un nombre fini de fonctions holomorphes dans V.

Une fonction $\varphi : A \longrightarrow \mathbf{C}$ est analytique si, pour tout point $a \in A$ la fonction φ est la restriction d'une fonction holomorphe sur un voisinage de a dans U.

Il est facile de faire de ces fonctions un faisceau $\mathcal{S}_A$ sur A, de la façon suivante.

Soit V un domaine contenu dans U. L'intersection $V \cap A$ est un ouvert de A pour la topologie induite, auquel on associe l'anneau des fonctions holomorphes sur $V \cap A$ au sens précédent. Si $a \in A$ est l'origine de $\mathbf{C}^n$ la fibre de ce faisceau est l'anneau (en fait l'algèbre sur $\mathbf{C}$) quotient de $\mathbf{C}\{z_1, \ldots, z_k\}$ par l'idéal I_a des germes en a de fonctions $\varphi : \mathbf{C}^k \longrightarrow \mathbf{C}$ qui s'annulent sur A. On ne considère que des faisceaux dont les fibres sont des anneaux réduits, c'est-à-dire sans éléments nilpotents.

Il semble que la terminologie courante soit que, tant que l'on considère des sous-ensembles dans $\mathbf{C}^k$ on parle d'"ensemble" analytique. Si l'on globalise, on parle d'"espace" analytique.

Ensuite, on globalise ainsi. Soit X un espace topologique séparé. En fait on peut tout aussi bien le supposer localement compact, car les conditions suivantes vont l'impliquer. Cartan suppose aussi que X est paracompact, ce qui nous convient très bien. Soit alors $\mathcal{C}_X$ le faisceau sur X des (germes de) fonctions continues à valeurs dans $\mathbf{C}$. On dit que X est muni d'une structure d'**espace annelé** si l'on se donne un sous-faisceau d'anneaux $\mathcal{S}_X \subset \mathcal{C}_X$. Serre dit que c'est Cartan qui a eu l'idée d'utiliser les faisceaux pour définir une structure obtenue par recollement de structures locales. Pour son élégance et sa fécondité, cette idée fut extraordinaire !

Définition 6.1 *Un* **espace analytique** *est un espace topologique X satisfaisant les conditions ci-dessus, muni d'un faisceau d'anneaux $\mathcal{S}_X$ qui est localement isomorphe au faisceau $\mathcal{S}_A$ d'un sous-ensemble analytique.*

Si X et Y sont des espaces analytiques, une application continue $\phi : X \longrightarrow Y$ est par définition un **morphisme analytique** si pour tout point $x \in X$ et tout germe φ de fonction analytique en $\phi(x) \in Y$ la composition $\varphi \circ \phi$ est un germe de fonction analytique en $x \in X$. Rappelons qu'un germe en $x \in X$ est représenté par une fonction définie sur un voisinage de x.

En termes de faisceaux, cela revient à dire que l'image réciproque $\phi^{-1}(\mathcal{S}_Y)$ par ϕ du faisceau structural $\mathcal{S}_Y$ est un sous-faisceau du faisceau $\mathcal{S}_X$.

7. Appendice 2 : Espaces normaux

Voici, dans les grandes lignes, ce qu'il faut savoir sur les espaces analytiques normaux, en mettant l'accent sur l'aspect topologique. La référence historique est, bien sûr, les Séminaires Cartan 1951-1952 et 1953-1954. Les livres de Grauert-Remmert [7] et de Lojasiewicz [8] sont très complèts, avec tout ce qu'il faut d'algèbre pour soutenir la topologie. Le livre de Grauert-Remmert est consacré aux espaces "globaux" et fait donc, comme son titre l'indique, grand usage des faisceaux ; tandis que le livre de Lojasiewicz fait une étude locale des ensembles analytiques, sans utiliser de faisceaux. J'ajoute qu'il faut un travail assez considérable pour obtenir les énoncés ci-dessous, à partir des définitions de nature algébrique.

Soit X un espace analytique, que nous supposerons connexe dans ce qui suit. Notons $X_0 \subset X$ l'ouvert dense des **points lisses** de X. Le fermé complémentaire $X \setminus X_0$ est **l'ensemble** $Sing(X)$ **des points singuliers** de X. C'est un sous-ensemble analytique.

Définition 7.1 *(C'est un théorème suivant le point de vue adopté)*

(1) L'espace X est **irréductible** *si X_0 est connexe.*

(2) L'espace X est **localement irréductible** *si tout point possède un système fondamental de voisinages $\{U_j\}_{j\in J}$ tel que $U_j \cap X_0$ est connexe pour tout $j \in J$.*

Un espace irréductible X est de "dimension pure". Cela signifie que la dimension ponctuelle $dim_x X$ en un point $x \in X$ est la même pour tout point x. Par définition $dim_x X$ est le maximum de la dimension des variétés lisses connexes qui sont au voisinage de x.

Une autre propriété intéressante des espaces localement irréductibles est la suivante :

Soit X un espace localement irréductible et soit $\varphi : X_0 \longrightarrow \mathbf{C}$ une fonction holomorphe. Alors φ est localement bornée au voisinage de $Sing(X)$ si et seulement si elle s'étend en une fonction continue sur $Sing(X)$. Voir [8] p.351.

Ceci nous permet de passer au espaces normaux. Voir [8] p.337.

Définition 7.2 *Un point $x \in X$ est dit* **normal** *si pour tout voisinage ouvert $x \in U \subset X$ toute fonction holomorphe et bornée $\varphi : U \cap X_0 \longrightarrow \mathbf{C}$ s'étend en une fonction* *holomorphe sur U.*

Un **espace X est normal** *s'il est normal en chacun de ses points.*

Un espace analytique normal est localement irréductible.

La propriété essentielle des espaces normaux que nous avons utilisée est la suivante : Si X est normal et si $\varphi : X \longrightarrow \mathbf{C}$ est continue sur X et analytique sur X_0 alors φ est analytique partout.

Dans la discussion de ces propriétés intervient la notion de sous-espace "thin" (mince). Pour ce qui nous concerne, cela signifie que, dans un espace normal, un sous-espace analytique strict ne sépare nulle part.

Pour les espaces analytiques, il y a aussi une opération de normalisation. Si l'on oublie les faisceaux et si l'on triangule, alors on obtient la normalisation des pseudo-variétés dont nous avons parlé plus haut. Une autre propriété intéressante est que, si l'on part d'un espace analytique localement irréductible sa normalisation est topologiquement un homéomorphisme. Dans ce cas, ce que l'on fait en normalisant est enrichir le faisceau structural.

Retour à l'uniformisation des fonctions algébriques

Revenons à l'hypersurface Σ_P et à l'application $\Pi_P : \Sigma_P \longrightarrow \mathbf{C}^n$ qui permet d'uniformiser la fonction algébrique associée au polynôme $P(z,w)$ définie au §5.1. En général, Σ_P n'est pas normale. Considérons donc la normalisation $\widehat{\Pi}_P : \widehat{\Sigma}_P \longrightarrow \Sigma_P$ de Σ_P. Sauf si $n = 1$, $\widehat{\Sigma}_P$ n'est pas nécessairement lisse. Ce n'est pas non plus une hypersurface en général. Cependant, la composition $\Pi_P \circ \widehat{\Pi}_P : \widehat{\Sigma}_P \longrightarrow \mathbf{C}^n$ est un morphisme fini entre espaces normaux (le but est même lisse). Cette composition satisfait donc toutes les affirmations de la conclusion du théorème de Cartan énoncé au §2. Si l'on peut dire, la composition $\Pi_P \circ \widehat{\Pi}_P : \widehat{\Sigma}_P \longrightarrow \mathbf{C}^n$ est ce que l'on peut faire de mieux pour uniformiser la fonction algébrique associée au polynôme $P(z,w)$.

Plus précisément, la fonction algébrique est uniformisée sur Σ_P par la projection π_w sur l'axe des w. Elle est aussi uniformisée sur $\widehat{\Sigma}_P$ par la composition $\pi_w \circ \widehat{\Pi}_P$. La structure normale sur $\widehat{\Sigma}_P$ assure que cette fonction est analytique, puisqu'elle est continue en chaque point et holomorphe aux points lisses.

Un exemple simple où l'hypersurface Σ_P n'est pas normale est fourni par les fonctions algébriques des polynômes

$$P(z_1, z_2, w) = w^c - z_1^a z_2^b$$

avec $c, a, b \geq 2$. La normalisation a une singularité isolée à l'origine. Topologiquement c'est le cône sur un espace lenticulaire. Ce sont les singularités de Hirzebruch.

Le résultat suivant fait le lien entre divers concepts dont nous avons parlé. Voir [7] fin du §8.3.

Théorème 7.1 *Soit X un espace analytique normal de dimension n. Alors X est* **localement** *la normalisation d'une hypersurface dans $\mathbf{C}^{n+1}$.*

L'hypersurface est définie par un polynôme de Weierstrass

$$w^d + a_1(z)w^{d-1} + \cdots + a_{d-1}(z)w + a_d(z)$$

où les $a_i(z)$ sont des fonctions holomorphes en $z = (z_1, \ldots, z_n)$. Pendant longtemps on a appelé “algébroïde” la fonction multiforme associée à un tel polynôme, à coefficients holomorphes et non pas polynomiaux. Un cas plus général consiste à admettre des coefficients avec dénominateurs (c'est-à-dire méromorphes).

Remerciements : Merci à Françoise Michel pour ses judicieux conseils. Mais je suis entièrement responsable des erreurs qui subsisteraient.

Dans les références, je ne donne que la liste des articles directement concernés par mon propos. Il y a plusieurs livres qui traitent des finite analytic mappings. Par exemple ceux de Robert Gunning. Et évidemment il y a un certain nombre d'articles. Il y a également plusieurs articles sur les spreads. Voir la bibliographie de l'article de Montesinos. Aujourd'hui, grâce à Internet et MathSciNet, il est facile de localiser les textes pour lesquels je n'ai pas fourni de référence explicite.

J'ajoute que le livre de Gottfried Barthel, Friedrich Hirzebruch et Thomas Höfer [1] contient (p.269–271) une excellente présentation d'une partie des sujets traités ici. On y voit très bien comment on peut exploiter le théorème principal du §1, pour construire des variétés ayant des propriétés particulières et/ou un groupe d'automorphismes donné.

Références

[1] G. Barthel, F. Hirzebruch, T. Höfer : *Geraden-konfigurationen und Algebraische Flächen.* Aspecte der Mathematik Vieweg (1987) 308p.

[2] H. Behnke, K. Stein : *Modifikation Komplexe Mannigfaltigkeiten und Riemannscher Gebiete.* Math. Ann. **124** (1951) 1–16.

[3] H. Cartan : *Sur les fonctions de plusieurs variables complexes : les espaces analytiques.* Congrès International des Mathématiciens Edinburgh (1958) 33–52.

[4] H. Cartan : *Prolongements d'espaces analytiques normaux.* Math. Ann. **136** (1958) 97–110.
Voir aussi : *Prolongement des espaces analytiques.* Séminaire Cartan 1957–1958, Exposé 11.

[5] R. Fox : *Covering Spaces with Singularities.* In *Algebraic Geometry and Topology. A symposium in honor of S. Lefschetz.* Princeton University Press (1957) 243–257.

[6] H. Grauert, R. Remmert : *Komplexe Räume*. Math. Ann. **136** (1958) 245–318.

[7] H. Grauert, R. Remmert : *Coherent Analytic Sheaves*. Grundlehren der Mathematischen Wissenschaften, Springer Verlag, 1984.

[8] S. Lojasiewicz : *Introduction to complex analytic geometry*. Birkhäuser Verlag, 1991.

[9] J. M. Montesinos-Amilibia : *Branched Coverings after Fox*. Bol. Soc. Mat. Mexicana **11** (2005) 19–64.

[10] V. Puiseux : *Recherches sur les fonctions algébriques*. Journal de mathématiques pures et appliquées **15** (1850) 365–480.

[11] Reinhold Remmert : *From Riemann Surfaces to Complex Spaces*. Séminaires et Congrès 3, Société mathématique de France (1998) 203–241.

[12] Karl Stein : *Analytische Zerlegungen komplexer Räume*. Math. Ann. **132** (1956) 63–93.

Algunas contribuciones matemáticas del Profesor José María Montesinos Amilibia

María Teresa Lozano

Departmento de Matemáticas
Universidad de Zaragoza
50009 Zaragoza, Spain

tlozano@unizar.es

1. Pretexto

Se me ha encomendado la tarea de explicar el trabajo matemático de José María Montesinos, que yo he aceptado con entusiasmo porque tengo el privilegio de colaborar con él desde hace más de treinta años. Considero que es imposible plasmar en unas pocas páginas el trabajo de José Mari, como muchos le llamamos, por su amplitud y complejidad. José Mari tiene un profundo conocimiento de muchos temas matemáticos. Así que he decidido fijarme en destacar su contribución a la topología de dimensión baja, sobre todo de nudos y 3-variedades, que ha liderado en España y de la que es figura destacada internacionalmente.

Pero antes de enunciar parte de sus resultados, no me resisto a dar unas pinceladas de nuestra historia y a dibujar un somero retrato recordando algunas de sus habilidades.

2. Contexto

La primera vez que oí hablar de José María Montesinos, fue en una oficina del emblemático y antiguo Bascom Hall de la Universidad de Wisconsin en Madison. En esa oficina teníamos mesa algunos matemáticos realizando una estancia postdoctoral, visitantes y colaboradores del Departamento de Matemáticas, que estaba ubicado en un moderno edificio, el Van Vleck Hall. Entonces era moderno porque estoy hablando de un día en 1977 y el edificio se había construido pocos años antes. Ese día llegó un joven matemático americano para realizar una corta estancia en esa Universidad, y al constatar que yo era española me preguntó si conocía a José María Montesinos, al que había conocido en Princeton: Un joven y valioso español que había demostrado importantes resultados en su tesis, relacionados con las últimas investigaciones del

Profesor Ralf Fox. Aquello me interesó porque en esos años yo estaba descubriendo la belleza de la topología de dimensión baja gracias a cursos de Peter Scott, Paul Melvin y Jim Cannon.

El recuerdo de aquella conversación volvió a mi mente un par de años más tarde, en el curso 78-79, cuando José María Montesinos vino a Zaragoza formando parte de un tribunal. Conseguimos atracarle para que nos diera una conferencia, casi improvisada, de un día para otro. Fue perfecta, como suelen ser sus charlas, con tiza, pizarra y bellos dibujos. No debió parecerle mal el interés de la audiencia, 5 ó 6 personas, porque cuando sacó una de las dos agregadurías (Sevilla y Zaragoza) que salieron en 1979, tuvimos la suerte de que eligiera Zaragoza. Durante su estancia allí (unos seis años) invitó a muchos topólogos amigos suyos que había conocido epistolarmente o en Princeton (Hilden, Fico (González-Acuña), Murasugi, Matsumoto, Kauffman, Przytycki, ...) a realizar cortas estancias. Con ellos trabajábamos, hacíamos excursiones (es conocida su afición al monte), y disfrutábamos de sus dotes de tenor entonando bellas canciones, casi siempre vascas, tras alguna agradable cena o comida.

Yo diría que su gran afición, aparte de las matemáticas, es la naturaleza, que es el museo al aire libre mas interesante que existe. De ella admira sus montañas y paisajes. Conoce el nombre de casi todos los árboles y minerales. Sus pesquisas en la naturaleza le han permitido descubrir inscripciones pétreas antiguas que esconden algunos rincones perdidos. Su interés teórico y práctico por la cristalografía le ha llevado a veces a llenar sus estanterías de magníficos cristales naturales, e incluso a generarlos. También ha realizado coloridos poliedros de papiroflexia. Pero además no los atesora, los regala generosamente. Yo tengo en mi despacho una piedra negra de gran densidad que pesa varios kilos, encontrada en la cima de una montaña oscense y que tiene apariencia de meteorito, aunque según su teoría tiene otro origen.

Jose Mari es dueño, por decision de la comunidad matemática, de tres valiosas posesiones. Hay tres conceptos matemáticos que han sido bautizados con su nombre:

1. "Montesinos knots and links". Son nudos y enlaces formados por ovillos racionales encadenados. Los estudió en [18] donde demuestra que la cubierta doble de la esfera S^3 ramificada sobre un nudo de esa clase es una variedad de Seifert. En la Figura 1 se muestran sus dibujos originales.

2. "Montesinos trick". Este truco realiza cirugías de Dehn en una 3-variedad que es cubierta doble de S^3 ramificada sobre un enlace mediante cierta modificación del enlace que introduce un ovillo racional.([20])

3. "Montesinos moves". Son transformaciones o jugadas de diagramas de nudos que utilizó por primera vez en su tesis doctoral. Se aplican a cubiertas irregulares ramificadas sobre enlaces (la monodromía envía meridianos a trasposiciones). Son modificaciones del enlace que dejan invariante la variedad cubriente. ([24] y [25]). La Figura 2 pertenece a [24].

Es el único matemático español que conozco con un libro en japonés. La traducción de su libro [23].

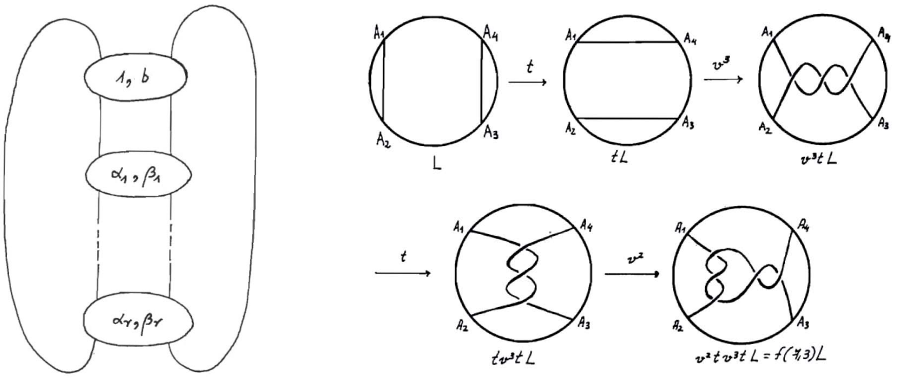


Figura 1: Nudo de Montesinos y ovillos

3. Texto

En lo que sigue, se agrupan los resultados por temas, tratando de cubrir los trabajos de Jose Mari que han recibido más citas, y los más recientes. Esto está contenido en menos del 20 % de su producción científica. Espero que esta muestra sirva para dar una idea de la importancia de los variados e ingeniosos resultados que ha obtenido hasta el momento, porque su actividad investigadora afortunadamente continua.

3.1. Cubiertas dobles de S3 ramificadas sobre enlaces

El primer trabajo sobre este tema [18], está escrito en español y publicado en el Boletín de la Sociedad Matemática Mexicana en 1973. En este trabajo se demuestra que las cubiertas ramificadas sobre ciertas clases de enlaces son variedades de Seifert. Una muestra de la importancia de este trabajo es que a pesar de estar escrito en español tiene más de cuarenta citas. Los enlaces de la clase que se define en el artículo son conocidos como enlaces de Montesinos tras la monografía [1] escrita por F. Bonahon y L. C. Siebenmann en 1979, que en 2010 vió su treintava edición.

Otro trabajo relacionado con el anterior y de gran influencia posterior es *Surgery on links and double branched covers of* S^3 [20], publicado en 1975 en el libro dedicado a la memoria de R. H. Fox. En este trabajo se estudia la relación entre las cubiertas de dos hojas de S^3 ramificadas sobre un enlace y las 3-variedades cerradas y orientables que se obtienen por cirugía en un enlace de S^3. Un resultado clave es que una 3-variedad M cerrada y orientable se obtiene por cirugía en un enlace fuertemente invertible (existe una involución que conserva la orientación de S^3 que induce en cada componente del enlace una involución con dos puntos fijos) si y solo si M es una cubierta doble de S^3 ramificada sobre un enlace. En este trabajo también se define y utiliza un concepto de cirugía generalizada que sustituye un conjunto de n toros

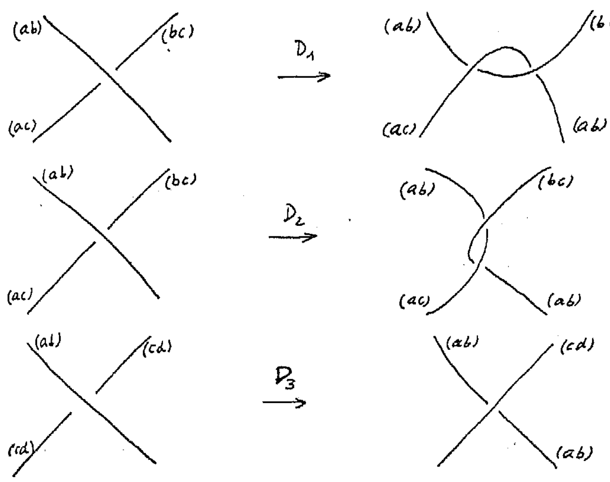


Figura 2: Jugadas de Montesinos

sólidos disjuntos en S^3 por una variedad de grafo cuyo borde está formado por n toros.

3.2. Las 3-variedades como cubiertas de S^3

Existen varios métodos para construir todas las 3-variedades cerradas y orientables, por ejemplo Diagramas de Heegaard, cirugía en un nudo o enlace en S^3, o cubiertas ramificadas sobre un enlace en S^3. Este último método fue introducido por Alexander en 1920. Montesinos ha contribuido a este método alcanzando los mejores resultados posibles en dos direcciones: número de hojas de la cubierta y lugar de ramificación más simple.

3.2.1 Cubiertas irregulares de tres hojas. En 1974 Montesinos publicó [19] en el Bull. AMS el teorema que asegura que toda 3-variedad cerrada y orientable es cubierta irregular de tres hojas de S^3 ramificada sobre un nudo. El mismo resultado fue encontrado casi simultáneamente por Hilden usando diferentes técnicas. Este es el mejor resultado posible en cuanto al número de hojas. En este tema son cruciales las llamadas jugadas de Montesinos en cubiertas de tres hojas irregulares que modifican la ramificación sin alterar la variedad cubriente y que Montesinos encontró en su tesis doctoral.

3.2.2 Nudos universales. Un enlace universal es aquel que tiene la propiedad de que cada 3-variedad cerrada y orientable es cubierta de S^3 ramificada sobre él. El concepto tiene origen en una comunicación privada fechada el 19 de febrero 1982 de Thurston dirigida a Montesinos. En ella se mostraba un complicado enlace con esa propiedad y se planteaban algunas cuestiones: ¿Existe un nudo universal? ¿Es el nudo Ocho 4_1 universal? ¿Es el enlace de Whitehead universal? Trabajabamos en este tema cuando Mike Hilden nos visitó en junio del mismo año. (Este fue el origen del equipo estable de investigación Hilden-Lozano-Montesinos con más de treinta años de fructífera colaboración.) En poco tiempo contestamos afirmativamente a todas las cuestiones planteadas por Thurston y encontramos nuevos resultados, [6], [8], [9]... . El nudo Ocho 4_1 es el nudo con menos cruces que puede ser universal. En colaboración también con Whitten se demostró que los anillos de Borromeo constituyen un enlace universal [5]. En este trabajo se introduce el concepto de grupo universal.

3.2.3 Grupo universal. [5]. Un grupo de isometrías hiperbólicas conservando la orientación que tiene como dominio fundamental un dodecaedro regular hiperbólico con ángulos diedrales de $90°$ y cuyo cociente es S^3 con los anillos de Borromeo como singularidad de orden 4, es el grupo universal U: Para cada 3-variedad cerrada y orientable M existe un subgrupo de índice finito G de U, tal que $M = H^3/G$. En otras palabras, M es un orbifold hiperbólico que es cubierta finita del orbifold hiperbólico $S^3 = H^3/U$. Este resultado traslada problemas de clasificación de 3-variedades en problemas de clasificación de subgrupos de índice finito de U. El estudio de los subgrupos de índice finito del grupo universal U, y el estudio de otros grupos universales ha sido abordado en otros trabajos en colaboración con Hilden, Lozano y otros colaboradores.

3.3. Variedad de caracteres

En el artículo [4] de J.M. Montesinos en colaboración con González-Acuña en 1993 se da una nueva demostración, larga pero elemental, del resultado de M. Culler and P. B. Shalen [3] de que la variedad de caracteres $X(G)$ de las representaciones de un grupo G en $SL(2, \mathbb{C})$ es un conjunto algebraico cerrado. Además en el trabajo se dan explícitamente un conjunto de polinomios que lo definen. En varios trabajos realizados en colaboración con Hilden y Lozano se calcula la componente excelente de la variedad de caracteres del grupo de algunas clases de nudos hiperbólicos: nudos de 2 puentes, nudos y enlaces periódicos, nudos con número de túnel 1... . También se obtienen nuevos polinomios invariantes asociados al nudo, los polinomios periféricos ([7], [16]) que definen la componente excelente de la variedad de caracteres y que están relacionados con el polinomio A definido por Cooper y Long in [2].

3.4. Orbifolds aritméticos

Un orbifold hiperbólico de dimension 3 es aritmético si es el cociente del espacio hiperbólico H^3 por la acción de un subgrupo aritmético de $SL(2, \mathbb{C})$, es decir un grupo de Klein aritmético. La caracterización y el estudio de estos grupos es un campo de interés porque proporciona subgrupos discretos de $SL(2, \mathbb{C})$ de covolumen finito, y por tanto orbifolds hiperbólicos, por procedimientos puramente aritméticos. El trabajo realizado en este tema por Montesinos y sus colaboradores incluye una caracterización en términos bastante elementales de esos subgrupos [11] y la obtención de todos los orbifolds aritméticos cuyo espacio subyacente es la esfera S^3 y cuya singularidad es el nudo Ocho [10], o los anillos de Borromeo [12], o un nudo hiperbólico racional [13]. Resultando en todos los casos estudiados que el número de tales orbifolds aritméticos es finito mientras que existen infinitos orbifolds hiperbólicos con el mismo espacio subyacente y el mismo conjunto singular.

3.5. Variedades cónicas geométricas

Las variedades cónicas geométricas aparecen como generalización natural de los orbifolds, permitiendo que el ángulo en torno a la singularidad tenga un valor que no sea divisor de $360°$. Se estudian así familias continuas de variedades cónicas que contienen a los orbifolds con el mismo conjunto singular como subconjunto discreto. Esta continuidad es útil para estudiar invariantes como el volumen y el invariante de Chern-Simon mediante fórmulas de integración. La contribución a este tema está contenida en los trabajos [14] y [15].

3.6. Dimensión 4

Los trabajos [21] y [22] son sobre maclas en la esfera de dimensión 4. Una macla consta de dos 2-nudos que se intersecan transversalmente en dos puntos. En [21] se define el concepto y se estudia su complemento. En [22] se estudian problemas de fibración en ese contexto dando nuevas demostraciones de resultados de Zeeman y Litherland.

3.7. Variedades de mosaicos

En el libro titulado *Classical tessellations and three-manifolds* [23] publicado en 1987 por Springer-Verlag se estudian las 3-variedades de posiciones de los mosaicos bidimensionales. Estos mosaicos son esféricos, planos o hiperbólicos. Los mosaicos han sido utilizados frecuentemente en arte: La Alhambra, Escher,.... En este libro se utilizan como introducción natural a un interesante tipo de 3-variedades, las variedades de Seifert. Por ejemplo, cada uno de los 17 posibles patrones de mosaicos en el plano da lugar a una 3-variedad, considerando una posición fija del mosaico como punto de partida, se traslada o gira y cada una de las nuevas posiciones define un punto de la variedad. En ocasiones dos posiciones corresponden al mismo punto porque los patrones coinciden. La variedad resultante está definida por un conjunto de invariantes que la clasifican. El libro, del que existe una versión previa en español, ha sido traducido al japonés.

3.8. Aplicaciones pseudo-periódicas y degeneración de Superficies de Riemann

El libro titulado *Pseudo-periodic maps and degeneration of Riemann surfaces* [17] publicado por Springer, escrito en colaboración con Matsumoto ha sido publicado en 2011, aunque la primera memoria fue escrita en 1991. El libro tiene dos partes. En la primera parte se completa el estudio iniciado por Nielsen en 1944 sobre aplicaciones pseudo-periódicas dando un conjunto completo de invariantes de la clase de conjugación de tales aplicaciones. La segunda parte contiene aplicaciones de los resultados de la primera parte a la teoría de degeneración de superficies de Riemann.

3.9. Formas cuadráticas

En los últimos años, y en su afán por entender a fondo la relación aritmeticidad-geometría, se está dedicando a estudiar la clasificación de formas cuadráticas enteras. Aunque el interés del estudio de estas formas cuadráticas se remonta al siglo XIX, no se conoce todavía un sistema completo de invariantes que las clasifique. Montesinos ([26], [27] y [28]) ha definido los conceptos de conmensurables, proyectivamente equivalentes y Bianchi equivalentes en el conjunto de formas cuadráticas racionales, los ha relacionado entre sí y con el concepto clásico de racionalmente equivalentes. En sus trabajos sobre el tema estudia invariantes para las correspondientes clases de equivalencia. La teoría de formas cuadráticas enteras tiene también un interés geométrico, porque algunas están relacionadas con modelos de geometrías no euclídeas. El grupo de automorfismos de la forma constituye el grupo de isometrías de la geometría correspondiente y sus subgrupos propiamente discontinuos definen orbifolds. Un bellísimo ejemplo de este hecho se muestra en [26].

Referencias

[1] F. Bonahon, L. C. Siebenmann: *New geometric splittings of clasical knots and the classification and symmetries of arborescent knots.* J. Knot Theory Ramifications, page 365, 1979.

[2] D. Cooper, D. D. Long: *Remarks on the A-polynomial of a knot.* J. Knot Theory Ramifications, 5(5) 609–628, 1996.

[3] M. Culler, P. B. Shalen: *Varieties of group representations and splittings of* 3*-manifolds.* Ann. of Math. (2), 117(1) 109–146, 1983.

[4] F. González-Acuña, J. M. Montesinos-Amilibia: *On the character variety of group representations in* $\mathrm{SL}(2,\mathbf{C})$ *and* $\mathrm{PSL}(2,\mathbf{C})$. Math. Z. 214(4) 627–652, 1993.

[5] H. M. Hilden, M. T. Lozano, J. M. Montesinos, W. C. Whitten: *On universal groups and three-manifolds.* Invent. Math. 87(3) 441–456, 1987.

[6] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *Universal knots.* Bull. Amer. Math. Soc. (N.S.) 8(3):449–450, 1983.

[7] H. M. Hilden, M. T. Lozano, J. M. Montesinos Amilibia: *Character varieties and peripheral polynomials of a class of knots.* J. Knot Theory Ramifications 12(8):1093–1130, 2003.

[8] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *The Whitehead link, the Borromean rings and the knot* 9_{46} *are universal.* Collect. Math. 34(1):19–28, 1983.

[9] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On knots that are universal.* Topology 24(4):499–504, 1985.

[10] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *The arithmeticity of the figure eight knot orbifolds. In* Topology '90 (Columbus, OH, 1990), volume 1 of *Ohio State Univ. Math. Res. Inst. Publ.*, pages 169–183. de Gruyter, Berlin, 1992.

[11] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *A characterization of arithmetic subgroups of* $\mathrm{SL}(2, \mathbf{R})$ *and* $\mathrm{SL}(2, \mathbf{C})$. Math. Nachr. 159, 245–270, 1992.

[12] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On the Borromean orbifolds: geometry and arithmetic.* In *Topology '90 (Columbus, OH, 1990)*, Ohio State Univ. Math. Res. Inst. Publ. (1) 133–167. de Gruyter, Berlin, 1992.

[13] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On the arithmetic* 2*-bridge knots and link orbifolds and a new knot invariant.* J. Knot Theory Ramifications 4(1) 81–114, 1995.

[14] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On volumes and Chern-Simons invariants of geometric* 3*-manifolds.* J. Math. Sci. Univ. Tokyo 3(3) 723–744, 1996.

[15] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *Volumes and Chern-Simons invariants of cyclic coverings over rational knots.* In *Topology and Teichmüller spaces (Katinkulta, 1995)*, 31–55, World Sci. Publ., River Edge, NJ, 1996.

[16] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *Peripheral polynomials of hyperbolic knots.* Topology Appl. 150(1-3) 267–288, 2005.

[17] Y. Matsumoto, J. M. Montesinos-Amilibia: *Pseudo-periodic maps and degeneration of Riemann surfaces.* Lecture Notes in Mathematics, vol. 2030, Springer, Heidelberg, 2011.

[18] J. M. Montesinos: *Variedades de Seifert que son recubridores cíclicos ramificados de dos hojas.* Bol. Soc. Mat. Mexicana (2) 18, 1–32, 1973.

[19] J. M. Montesinos: *A representation of closed orientable* 3*-manifolds as* 3*-fold branched coverings of* S^3. Bull. Amer. Math. Soc. 80, 845–846, 1974.

[20] J. M. Montesinos: *Surgery on links and double branched covers of* S^3. In *Knots, groups, and* 3*-manifolds (Papers dedicated to the memory of R. H. Fox)*, pages 227–259. Ann. of Math. Studies, No. 84. Princeton Univ. Press, Princeton, NJ, 1975.

[21] J. M. Montesinos: *On twins in the four-sphere, I.* Quart. J. Math. Oxford Ser. (2) 34(134) 171–199, 1983.

[22] J. M. Montesinos: *On twins in the four-sphere, II.* Quart. J. Math. Oxford Ser. (2) 35(137) 73–83, 1984.

[23] J. M. Montesinos: *Classical tessellations and three-manifolds.* Universitext. Springer-Verlag, Berlin, 1987.

[24] J. M. Montesinos Amilibia: *A note on a theorem of Alexander.* Rev. Mat. Hisp.-Amer. (4) 32, 167–187, 1972.

[25] J. M. Montesinos Amilibia: *Reduction of the Poincaré conjecture to other geometric conjectures.* Rev. Mat. Hisp.-Amer. (4) 32, 33–51, 1972.

[26] J. M. Montesinos-Amilibia: *On integral quadratic forms having commensurable groups of automorphisms.* Hiroshima Math. J. 43(3) 371–411, 2013.

[27] J. M. Montesinos-Amilibia: *Addendum to "On integral quadratic forms having commensurable groups of automorphisms" [MR 3161323].* Hiroshima Math. J. 44(3) 341–350, 2014.

[28] J. M. Montesinos Amilibia: *On binary integral quadratic forms having isometric groups of automorphisms.* Bol. Soc. Mat. Mex. (3) 20(2) 199–236, 2014.

José María Montesinos Amilibia
Biographical Sketches

Sebastià Xambó Descamps *

Department MA2
Universidad Politécnica de Cataluña
08034 Barcelona, Spain
sebastia.xambo@upc.edu

To José María Montesinos Amilibia
in appreciation for his mathematics, wisdom, and friendship.

My interest in the work and life of José María Montesinos Amilibia (JMMA) was first kindled when he asked me whether I could draw the plane unfolding of a polyhedron that *tesselated the Euclidean and* (in a way, v. infra) *the hyperbolic* 3-*space*. Seemingly it had just been discovered in a joint work with Hugh Hilden and María Teresa Lozano Imízcoz (MTLI). At that time I was in the Algebra Department (UCM[1]) and took the exercise in my trip to Barcelona on the coming weekend. The data he gave me were the coordinates of the vertexes and on next Monday I was happy to present him copies of a plane unfolding ready to cut, fold and paste and also a couple of mounted pieces (Figures 1 and 2).

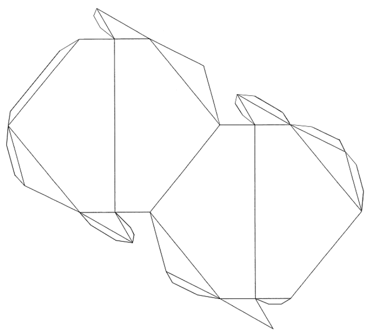


Figure 1: Plane development of the eupolytope.

The result was quite amazing, particularly because the means that then were at my disposal were very rudimentary. If it is true that I had a lot of fun doing it, I do not think that it brought anything to JMMA that he did not already know. Later he told me that he had filled one of the paper models with methacrylate, so that he got a solid transparent model. The vertexes' coordinates can be found in [76], the first paper in which this polyhedron is described and studied in detail. A surprising fact is that *if the dihedral angle at the two longest edges, which is* $120°$, *is decreased to any non-negative value, then the new polyhedron tesselates the hyperbolic space*. Taking into account the meaning of the Greek prefix 'eu',

*Partially supported by the ArbolMat project (http://www.arbolmat.com/).

[1] For the meaning of acronyms, see the table at the end.

which in scientific coinages connotes ‘good, true, genuine’ (the etymologies of ‘euphony’ or ‘eulogy’ are good examples), I suggested to call it *eupolytope*, or *eupolyhedron*, but I have the impression that for some reason it did not catch, possibly because in his universe all polyhedra would deserve that prefix.

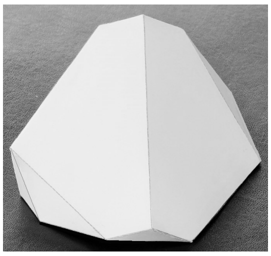
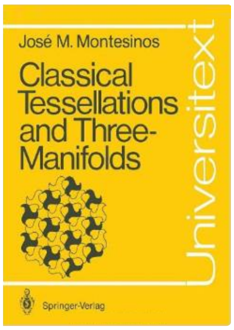


Figure 2: Paper model of the eupolyhedron.

Another precious spark that came a little later was produced in relation to the volume of mathematical contributions in honor of José Javier Etayo Miqueo (UCM, 1994). The paper I wrote, [149], owes a great deal to the questions posed at the start by JMMA (What is the spectrum of a ring? What is it useful for? Is it possible to do without such an abstract concept? . . .) and to the discussions with him all along. In particular I tried hard to show in what ways the spectrum of a ring is naturally connected to the physical notion of spectrum, that is, the set of wavelengths of radiation emitted by (say) atoms. It is not clear to me whether he thought that his questions were answered in that work, but my impression that it was worth while to try still lives in me. Different connections between algebra/analysis and geometry/physics were considered and also the deep analogies between them.

Twenty years later, in the context of the *Árbol de las Matemáticas*, I was involved in the preparation of his profile [153]. Basically it amounts to a mention of a sample of his merits, including a few of his outstanding works, and a sketchy academic biography. It was published in September 2014 and because of that experience, I took as a great honor the invitation that José Manuel Rodríguez Sanjurjo sent me to participate in the homage volume and conference that was beeing planned by the UCM Geometry Department. I could accept because it was clear that MTLI would be in charge of reporting about JMMA main scientific contributions, a job that can be properly done only by a close and sustained collaborator as she is.

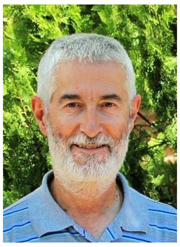


Since *by their works ye shall know them*, the first task I thought I had to undertake was to compile a complete list of JMMA’s works. The result can be found in the References at the end, which are grouped in four categories: Books ([1]-[5]), Memoirs ([6]-[10]), Papers ([11]-[124]) and Articles ([125]-[134]). The section External references ([135]-[154]) contains works not authored by JMMA that are cited in these notes. Most of them are closely related to his work. The entries in each category are ordered by the year of publication.

I also tried to assemble copies of all those references, a task that is practically complete at the time of this writing thanks to the generous help provided by JMMA and MTLI. I regard these materials as the bedrock on which any perspective overview of JMMA scientific profile must be based. My gratitude goes also to Ángel Montesinos Amilibia, for his hospitality in two visits at the home he and José María share in Cebreros (Ávila).

Origins

José María Montesinos Amilibia was born in San Sebastián, the capital of Guipúzcoa (Basque Country, Spain) on 13 November 1944. His father, Lorenzo Montesinos, was a school teacher. In 1938 he was wounded in Teruel, where he fought in the Spanish Civil War enlisted in the "5ª Brigada Navarra". It was a severe bullet injury to one of his legs and for his recovery he chose what he thought was his best option: having made many good friends among the Basque soldiers serving in the brigade, he asked to be carried to a hospital in San Sebastián. There a wise doctor managed, after several delicate operations, to save his leg. He also met Victoria Amilibia, whom he married soon. This union was blessed with five offspring: María Victoria "Mariví", Ángel, José María, Antonio and Coro "Corito". Of them, Ángel and José María were to become mathematicians (although originally Ángel is physicist), while the others pursued different professions.



Á. Montesinos

The basque family name Amilibia has a close linguistic tie to the English family name Clifford. Indeed, if we split both in two parts, 'amil-ibia' and 'clif(f)-ford', the connection is clearly revealed because the meaning of the Basque root 'amil' is close to the English meaning of 'cliff' ('amildegi' in Basque), and similarly with 'ibi' and 'ford'. This was recently told by JMMA when, on launching a symbolic computation program on geometric algebra, the name and the picture of Clifford appeared in the screen.

On the other hand, the family name Montesinos is connected with Montesinho, the Portuguese version of the Montesino in the Spanish Zamora province, the birthplace of Lorenzo. There is, for instance, a Montesinho National Park just below the northern Portuguese border and not far from Bragança, Zamora's sister city. The fact that JMMA dedicated his landmark work [3] to his parents is a faithful and timely expression of the high esteem and veneration which he felt for them.

Schooling

Through his primary and secondary eduction, JMMA was at the same time a cheerful and lively boy and a mind that took his studies very seriously. He was schooled in the Colegio Sagrado Corazón, in San Sebastián (Mundaiz street, 30). He liked all subjects, except History, and at the same time he engaged in all sorts of playful activities.

The subject he liked best was Chemistry, so much so that he reached a real expert level before entering the university. Then came Mathematics and Physics, but also Grammar, a discipline that he has cultivated all along in different guises, particularly linguistics and etymology. One mark of his cognitive endowment was already apparent in those studies: his unlimited persistence in trying to reach a full understanding, in

Colegio Sagrado Corazón.

his own terms, of the objects and processes that interested him. He had a maturity that was far above the average for his age.

University

The determination of JMMA at the end of the secondary education was to study Chemistry, a subject that, as already indicated, had interested him in high school well above the others and about which he had an extensive knowledge.

At that time, this plan required to enroll in a Science Faculty for a five-year program. These programs were taught in several universities in Spain and consisted of a first general common course whose core was formed by one subject for each of the possible major specialties that could be chosen at the beginning of the second year: Mathematics, Physics, Chemistry, Biology and Geology. The first course had a selective character, in the sense that it was required to have passed its main subjects in order to begin one of the specialties in the second year.

JMMA chose the Science Faculty of the University of Barcelona, where he entered in the Fall of 1962. The experiences of that course had a major impact on his career, for at the end he decided to pursue Mathematics. The first strand for that change arose from a circumstantial contingency. Students were assigned to one of several groups, but for some reason he did not know his group until the end of October. Thus meanwhile he attended the classes of the first group, where Mathematics was in charge of Rafael Mallol Balmaña, a young teacher that had just read his doctoral thesis under the direction of Enrique Linés Escardó. JMMA remembers those lectures as truly enlightening about the nature of Mathematics and its ways, especially because of the crystal clear construction of the basic number systems. Thereafter he *knew* how to approach the study of Mathematics and has been grateful ever since for having learned those basic principles at that moment from an excellent teacher.

The second factor seems to have been that he judged himself, after some minor accident in the Chemistry laboratory, as not having the proper level of manual dexterity. And the third reason was that he eventually was assigned to the group in which Mathematics was being taught by Juan Augé Farreras, a former student of Ricardo San Juan Llosá. He found the learning of Mathematics to be rather easy-going and at the end he got the highest mark. The three strands tied him in a dilemma: to begin a Chemistry career, as initially planned, for which he was well prepared in terms of knowledge but rather insecure about his experimental skills, or to switch to Mathematics, which he now understood and appreciated but for which he did not perceive himself as having any special aptitude even after having been bestowed with the highest qualification. So he approached Professor Augé for advice, who answered with his legendary phlegmatic poise: "Mr. Montesinos, go for Mathematics".

In that year, he was also highly interested in Biology, taught by Professor Enrique Gadea Buisán, and in Geology, taught by Professor José Ramón Bataller Calatayud until his sudden death shortly before Christmas. His Geology studies were eased by his knowledge of Chemistry and he has kept an interest in it ever since, particularly intense in the area of Crystallography, in part because of its deep ties with Geometry and Topology.

Clock tower, UB.

Most of the intense studies of that year were carried out, completely undisturbed, in the clock tower of the historical building of the University of Barcelona. He found his way there and apparently nobody ever noticed. He deemed that the noises emitted by the clock machinery and by the bells on top were agreeable reinforcers of his work schedule. Perhaps it was also an ideal spot from where to play some stealth innocent tricks to the people going about their affairs several stories below.

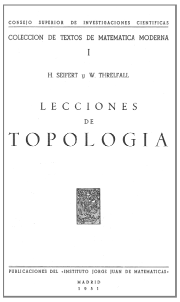
For the four remaining years of Mathematics studies he moved to the UCM. He has good memories of the excellent lectures on Analytic Geometry by Professor Francisco Botella Raduán. Joaquín Arregui Fernández taught him Algebra and Topology. He also organized a seminar on Topological groups, which JMMA attended. He found that the lectures of Germán Ancochea Quevedo on Differential Geometry were great, and also those of Enrique Outerelo Domínguez on General Topology and later in the doctoral courses. He is grateful to José Luis Pinilla for his advice about how to teach problem solving in Geometry, and to Juan Fontanillas Royes and Feliciana Serrano for a similar reason in the case of Topology.

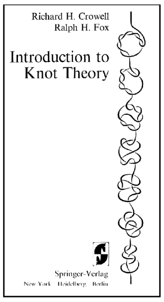
Early research

The first results of JMMA in research were in the period 1969-1971. They had a marked autodidact character and were obtained under rather adverse circumstances.

As most university students of that time, for his military service he chose the "university militias" option, which consisted in two four-months training periods in a military Summer camp, El Robledo (La Granja, Segovia) in his case, followed by a four-months period of service as second lieutenant (alférez) in a military base, which in his case was Cerro Muriano (Córdoba).

When he was called to that service, it was for an immediate incorporation and in a hurry he picked the (Spanish translation of the) treatise [135] and the text [140].



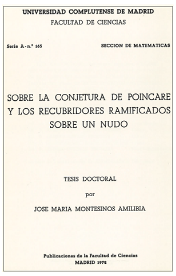


The circumstances in Cerro Muriano, and his cunning ways of dealing with them (the shadow of a holm oak as a regular office instead of attending the military instruction on the tasks that he would be later assigned), allowed him to study those texts in depth, so that he had a deep knowledge and understanding of them when he came back to Madrid at the end of 1969. But in January 1970 he felt sick, with high fevers, and it turned out that he had tuberculosis. It took one year for his full recovery, mostly spent at the Valdelatas Hospital (Madrid). Under these hard circumstances, he managed nevertheless to seize enough time and strength to carry on the research he had begun in Córdoba.

A key move after studying [140] was getting a copy of [136] (it is likely that the intercessor to bring it to Valdelatas was Professor Botella). The book is the proceedings volume of the 1961 meeting on *Topology of* 3*-manifolds and related topics* held a the Mathematical Institute of the University of Georgia under the auspices of the Office of Naval Research and the National Science Foundation. It is a superb and awesome book that "presents summaries and full-length reports of the Institute's five seminars, which covered decompositions and subsets of E^3; n-manifolds; knot theory; the Poincaré conjecture; and periodic maps and isotopies".

R. Fox.

One of the heroes of that meeting was Ralph Hartzler Fox, a Princeton Professor in his late forties. He is the single author of four papers included in [136], three among the six in the knot theory section (that include the survey [137]) and one among the seven in the Poincaré conjecture section, namely the three-page summary [139] about *constructions of simply connected* 3*-manifolds*. There is a fourth paper in collaboration with O. G. Harrold in the knot theory section and he was about to publish the first version of [140], the masterful memoir (actually a more elaborate version of [137]) with which JMMA was already thoroughly familiar at the outset of his illness. The focus of JMMA inquiry was [137], where the mood is that some simply connected branched coverings of S^3 might be counterexamples to the Poincaré conjecture.



With his characteristic persistence, JMMA introduced new techniques that allowed him to solve Fox's problem, that is, to show that all 3-manifolds produced with Fox's approach were homeomorphic to S^3. Loosely speaking, the core of the method relies on modifying the ramification of a branched cover of S^3 (without modifying the covering manifold) by means of what came to be known as *Montesinos' moves*, a procedure that has had many interesting applications ever since (v. [154]). With this, he could complete his doctoral thesis [1] toward the end of the 1970 Summer and read it the following year.

Consolidation

The completion of the thesis marked the beginning of a five-year period that was decisive to consolidate him as a researcher. The value of his work and ideas was soon recognized by first rank experts and as a consequence he gradually got in contact with them. It is enough to check the reviews of the papers published in that period, and who signed them, to realize that he was accepted (and perhaps even feared by some) as a new star in the field of low-dimensional topology.

The first important contact was with R. H. Fox, to whom he sent his doctoral memoir and the two subsequent papers he had published in volume 32 of the RMHA, [12] and [13]. The focus of [13] was the solution of Fox's problem, while [12] included two conjectures (A and B) that implied the Poincaré conjecture and that the Poincaré conjecture implied B. Fox reported about these works in his seminar at Princeton and then published the note [141], which cites the three mentioned works and in part is a summary of the thesis [1]. In addition, Fox proves a theorem that shows that conjecture A is false as stated and after some discussion proposes a modified form A' of A (adding a hypothesis of simple-connectednes) and a conjecture B' that has (at least to the eyes of a non expert) more of a new conjecture than a modified form of B. Jointly, A' and B' are equivalent to the Poincaré conjecture.

The next milestone was the publication of [14] and [15]. Of the latter, for example, the reviewer (H. E. Debrunner) says that it "is a fine piece of geometry, being specified throughout with interesting examples". It is an early recognition of a steady characteristic trait in his research: the balanced dialectical counterpoint between deep conceptual thinking and well chosen examples presented with the delicacy of a jeweler.

The doors were being opened. Laurent Siebenmann invited him to impart a one-month seminar in Paris mainly focussed on the last cited paper. Since then, the knots introduced there are known a *Montesinos' knots*. He also started a collaborative correspondence with Hugh M. Hilden, Joan S. Birman and Francisco J. González-Acuña. The collaboration with Hilden, often involving other coauthors (particularly MTLI), has endured unabated since the first joint paper in 1976 with an average of one paper per year. The collaboration with F. J. González-Acuña produced over half a dozen papers in the period 1978-1993, while Birman coauthored just two titles (in 1976 and 1980, the first also signed by González-Acuña).

L. Siebenmann, H. M. Hilden, J. B. Birman, and F. J. González-Acuña

JMMA had dreamt about working with Fox in Princeton, but sadly Fox died after his sixtieth birthday (1973). Part of the dream, however, came about to be true in other ways. The start was an invitation of Birman to publish a paper in the volume that was being planned in the memory of Fox [142]. The result was [20], which in the volume comes just after John Milnor's contribution. The technique introduced in that paper is known as *Montesinos trick* and it turned out to be very productive. In view of the quality and height of his inquiries, Birman proposed him to visit the Princeton IAS. He was accepted (by Milnor) and during the two-year stay (1976-1978) he met a good many of the geometric topologists, some working in high dimensions (like Milnor, Siebenmann, and R. D. Edwards), others in low dimension (like William Thurston, Yukio Matsumoto, Cameron Gordon, Andrew Casson and Ronald Fintushel), and with occasional visits by Robion Kirby and R. H. Bing. In JMMA's own words, "the solution of the Poincaré conjecture in dimension 4 was cooked there; we all attended Thurston's lectures; I was fortunate enough to absorb all this and by 1978 I was a new man".

J. Milnor, W. Thurston, Y. Matsumoto, C. Gordon, A. Casson, R. Kirby.

Full Professor

Upon returning to Spain, JMMA won a position at the University of Zaragoza, first as an Associate Professor (1979-1981) and then as a Full Professor (1981-1986). Officially, the term 1981-1982 he was a Full Professor at the Universidad Autónoma de Madrid, but actually it was arranged that he could continue in Zaragoza. The full year 1985 he was on leave with a position at the MSRI. Finally, since 1986 he is Full Professor at the UCM, where he holds the chair of Analytic Geometry and Topology.

MTLI

He arrived at Zaragoza just when MTLI had returned from a two-term stay as an Honorary Fellow at the University or Wisconsin-Madison. She had read her doctoral thesis in 1974, published a number of papers in K-theory, and raised a family with her husband Julio Abad, a theoretical physicist. Those two academic terms were somehow a slightly delayed postdoc that had the effect, basically due to the courses she took from Peter Scott, Paul Melvin and Jim Cannon, of getting her involved in low dimensional topology, and it was there and in this connection that she heard about JMMA for the first time. The net effect was (cf. [154] about details of how neatly it worked out) that the mathematicians in Zaragoza welcomed warmly and enthusiastically the arrival of JMMA. In the first place, MTLI herself, who was ready to appreciate his thinking and contribute at the highest level with her own, but also Antonio Plans, José Luis Viviente, Elena Martín, Álvaro Rodés and Andrés Reyes.

JMMA taking possession of his first professorship.

JMMA was happy with the atmosphere he found and reciprocated by giving everything he could to the group and to each individual member. In addition to his own lectures, he invited many prominent researchers to Zaragoza, often several times, sometimes on sabbatical leave (like Hugh Hilden, Francisco González-Acuña, Louis Kauffman, Lee Rudolf, Hugh Morton, Kunio Murasugi, Jozef Przytycki, Claude Weber, . . .), and this was a major contribution to the international recognition of the department. In Kauffman's book [146], for example, we read in the introduction:

> These notes on the theory of knots comprise an expanded version of a seminar held in the Departamento de Geometría y Topología at the University of Zaragoza, Zaragoza, Spain during the winter of 1984. Due to the *supernatural enthusiasm and persistence* of the members of the seminar, the author was given the energy to record a (we believe!) careful set of notes, and to relish the process.

This wonderful book is dedicated to the memory of Andrés Reyes, who died in an unfortunate car accident in 1984 in his early thirties. Incidentally, JMMA had been a mentor of Andrés and his book [3] is dedicated to his parents, as said before, but also to the memory of "mi amigo Andrés".

In all those years (1979 to today), the research interests of JMMA have diversified and become ever wider: manifolds of dimension 4, existence of universal knots and groups, arithmetic groups (the last two in collaboration with MTLI and H. M. Hilden), or open manifolds and wild knots. At present he is also dealing with questions concerning the degeneration of geometries (in collaboration with MTLI) and the automorphism group of quadratic forms.

Very often his works get excellent reviews and have good citation records. Already in 1985, at the end of his period in Zaragoza, with two thirds of his production still in the future, the treatise [144] cites a dozen works for wich he is a single author and seven more for which he is a coauthor (one with Birman, three with González-Acuña and three with Hilden), only below Fox (29 citations) and Murasugi (30 citations).

One important aspect of JMMA's academic life has been advising students for their doctoral research. The following list summarizes the information about those he has advised: the year in which it was read, the name of the student, the title of the thesis and the university in which it was ascribed.

1979

- Lucía Contreras Caballero: *Esferas homológicas* (UCM)
- José Manuel Rodríguez Sanjurjo: *Teoría de la forma* (UCM)

1984

- Valerio Chumillas Checa: *Cubiertas dihédricas* (UCM)
- Antonio Félix Costa González: *Representación de 3-variedades mediante cubiertas* (UCM)
- Carmen Safont: *Sobre cubiertas ramificadas* (UZ)
- Leopoldo Villarreal Sáez de Urabain: *Recubridores localmente cíclicos* (UCM)

1990

- Milagros Izquierdo Barrios: *Estudio de subgrupos de grupos de caleidoscopios no euclídeos que son grupos de superficies* (UZ, codirigida por MTLI)

1994

- Alberto Borobia Vizmanos: *Matrices no negativas* (UNED)

1998

- Eva Suárez Peiró: *Poliedros de Dirichlet de* 3*-variedades cónicas y sus deformaciones* (UCM)

2006

- Rubén Vigara Benito: *Representación de* 3*-variedades por esferas de Dehn rellenantes* (UNED)

Collected works

Looking at JMMA's work so far, we first notice that he is the single author of more than half of the publications bearing his name. As we can see in the references, this includes four books (among five), the five memoirs, the ten articles, and over fifty papers. If we take into account the extension, this represents well over sixty percent of a total of nearly 3300 pages. In fact, if we imagine all the materials arranged in volumes, we could easily arrive at six quite large books of which two would be devoted to the publications with at least one coauthor.

The creative power required for these accomplishments is apparent through all his career. It can be seen in a most pure state in his doctoral thesis, which looks as if it blossomed out of a few lines from [139] (our emphasis):

> In 1920, Alexander showed [Note on Riemann spaces, BAMS 26, 370-372] that any orientable n-manifold Σ is a branched covering of S^n, and that, furthermore, when $n = 3$, the branching may be assumed to take place over a tame link L in S^3 and to have branching index ≤ 2 everywhere on Σ. Alexander makes this last statement without proof, but if follows from results obtained last century by **Clifford** [On the canonical form and dissection of a Riemann's surface, PLMS 8 (1877), 292-304]. Alexander's theorem, therefore, makes it reasonable to search among the branched coverings of S^3 for simply connected 3-manifolds. It turns out that one can find bushel baskets of them this way. *No doubt most of them, and **possibly all**, are actually S^3, but I have never had the patience to verify this except for the one simple example of the* 3*-fold irregular branched covering of the trefoil.*

Further witnesses for this view, and also for the originality and independence of his thinking, are the reviews of his early papers (like [15], already commented before, [17], or [20]) and by the fact that they count among the best cited in the field. We may include on this count the book [3]. As declared by the referee (R. Fenn, but my emphasis):

> Incidentally, for the benefit of Hispanophiles, this book produces photographic evidence ***once and for all*** *that all 17 plane symmetry patterns appear in the Alhambra.*

But JMMA has carried out (and carries) collaborative work with a number of researchers. From the well over two thousand pages of the 111 papers in the references, about 55 percent belong to works with at least one coauthor. Among the coauthors, there are two that stand out far above all others: MTLI and Hugh Hilden. They appear as coauthors of about 36 and 30 percent of those pages, respectively (counting papers, the order is the other way around: 34 and 40, respectively). Actually, the coalition HLM (Hilden, Lozano, and Montesinos) has signed 28 papers that in pages add up to 28 percent (see 40-42, 47, 48, 55, 60, 66-69, 71 75-79, 81, 83-85, 92, 94, 98, 100, 106, 108, 114]). Since these are being considered in [154], here we will just note two comments on the significance of a few of them, and in particular [56] (one of the two papers of JMMA bearing Whitten's signature, the other being [52]).

In the historical notes to the last chapter of [148], a chapter devoted to geometric orbifolds (in Spanish, JMMA calls them *orbificies* when seen as topological objects and *caleidoscopios* when they have geometric structure), the papers [66] and [67] are cited, together with papers by Weber-Seifert (1934), Meyerhoff (1985), and Adams (1992), "for some interesting examples". And then we find the following paragraph:

> It is an interesting fact due to Thurston that every closed orientable 3-manifold has a hyperbolic orbifold structure. In fact, every closed orientable 3-manifold is an orbifold covering space of the hyperbolic orbifold in Example 5 [based on a regular hyperbolic dodecahedron, as explained on page 703]

and as a reference the author gives [56]. This paper was reviewed by V. G. Turaev, giving a clear picture of that example, and stating that "this theorem offers *a new approach to the Poincaré conjecture*" (my emphasis), which perhaps may still be perceived as an opportunity for those that are not completely blinded by Grigory Perelman's great success early in the current century.

Similarly, in [147], and again in relation to orbifolds, we find that (my emphasis)

> A classification of simply-connected orientable three-dimensional compact orbifolds admitting geometric structures other than hyperbolic can be found in Dunbar [1988]. In this connection, the following *quite amazing result* of Hilden, Lozano, Montesinos and Whitten [1987] should be mentioned: there are discrete groups Γ of motions of the space Л^3 (called universal) such that any orientable three-dimensional compact manifold is homeomorphic to $\text{Л}^3/\Delta$, where Δ is an appropriate subgroup of finite index in Γ (whose action is not necessarily free).

The 1987 cited paper is again [56].

Going down our (quantitative) collaboration scale, we find Francisco González-Acuña (alias Fico), on one hand, and Débora María Tejada and Margarita María Toro Villegas (TT), on the other. Fico coauthored seven papers in the period 1978-1993 (77 pages): [27, 35, 36, 39, 70] as a sole coauthor (52 pages), and [23] and [54] (25 pages), that are also signed by Birman and Boileau, respectively. The paper [27], published in AM, exhibits 2-knots that have infinitely many ends and so it answers affirmatively the last question posed by Fox in [138] (Problem 40). In paper [39], a quasiaspherical n-knot K in S^{n+2} is constructed having infinitely many ends, thus disproving a conjecture of J. G. Ratcliffe. Paper [70] is outstanding for its technical virtuosity in providing an 'elementary' deduction of the equations defining the character variety of group representations in $SL(2,\mathbb{C})$. JMMA's esteem for Fico can be appreciated at length in [129], and also in the dedication of [117] ("Dedicado con agradecimiento y afecto a Fico González Acuña en su septuagésimo cumpleaños").

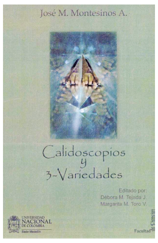
D. Tejada and M. Toro.

The researchers TT appear as coauthors in the period 2004-2012: they sign the six papers [95, 96, 99, 101, 110, 111] (124 pages), jointly with Hilden, and the two papers [104, 105] (39 pages), which are also signed by G. Brumfiel, MTLI, Ramírez-Losada, and H. Short. These works offer interesting results on a number of topics, but appear as somewhat technical to the non expert. An interesting exception may be [110], for we read in its summary:

> It is well known that there are 17 crystallographic groups that determine the possible tesselations of the Euclidean plane. We approach them from an unusual point of view. Corresponding to each crystallographic group there is an orbifold. We show how to think of the orbifolds as artifacts that serve to create tessellations.

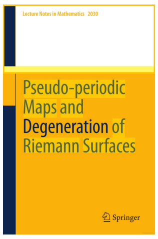
The collaboration of TT with JMMA began with the visits that he made to Medellín (Colombia) in 1995 and 1997. The invitations came from the "Grupo de

Investigación en Matemáticas" that TT had started. One early and important fruit of the lectures and courses that he taught there was the book [4], which roughly speaking is an introduction to the theory of orbifolds, a notion introduced by THURSTON in his Princeton lectures (1976-1977).



In the preface of the book, after explaining the importance of the concept of orbifold, JMMA says: "... and so, when I was invited to give a course of lectures in Medellín, I went with the idea of popularizing it". He found an enthusiastic group of participants, including TT, who took up the job of writing and polishing the text using JMMA's notes, and the students, that drew the figures with great care (there are dozens of them in numbered series and a great many that appear interspersed with the text with no numbering).

It is a pity that this book is not yet translated into English and that thus it is perhaps not as known as it deserves.



Of the remaining coauthors, YUKIO MATSUMOTO deserves a special consideration. He is the coauthor of papers [63, 74] and of the book [5]. Paper [63] provides a detailed proof of a statement on orbifolds phrased by THURSTON in his Princeton lectures that is closely related to his geometrization conjecture [150]. On the other hand, the writing of the book [5] was essentially "completed in December 1991, and some remaining additional parts in January, 1992" (2009 addenda to the Preface). The main reason for the nearly twenty years delay in its publishing was "the author's inability to use Tex" (*ibid.*). Because of this, the paper [74] now appears as a harbinger of the book. On the whole, the main points are the classification of the topological types of degenerate central fibers of holomorphic families of closed Riemann surfaces of genus $g \geq 2$ over the unit disc and the extension of NIELSEN's invariants to a complete set of conjugacy invariants for the pseudo-periodic homeomorphisms of negative twist.

Taking possession of his numerary seat at the RAC (1990). Left: with his father, Lorenzo. Right: reading his discourse.

A substantial part of the scientific and academic life of JMMA has revolved around the RAC. After taking the position as a Full Professor at the UCM, he was invited to give a lecture at the RAC. He chose the topic *Orbifolds in la Alhambra* (Caleidoscopios en la Alhambra) and subsequently he published it as the RAC memoir [8]. It is the year in which the book *Classical Tesselations and Three-Manifolds* [3] and the Inventiones paper *On universal groups and three-manifolds* [56] were published. The memoir is remarkable in many ways. In the epilog, he refers to the results in [56] and indicates what we have already commented before, namely, that they provide a new (geometric) tool to attack the Poincaré conjecture. Appendix

A is devoted to describe in detail a set of color pictures of mosaics in la Alhambra, thereby showing that all 17 plane crystallographic groups were known to the Nazari artists. Interesting details about who and how the different forms were discoverd, with due regard to the last one to be found (the elusive $D\bar{3}\bar{3}\bar{3}$) is provided in appendix B.

Although JMMA had been nominated as a RAC corresponding member in 1986, his next memoir published by the RAC is [9], which contains the discourse he read on the occasion (1990) of his accepting the nomination as a numerary member and a semblance piece (Discurso de Contestación) by José Javier Etayo Miqueo. As remarked by Etayo, it was the first time that Topology as such was incorporated in the RAC compass. In that solemn moment, JMMA summarized his intentions as follows (*ibid.*, page 9; emphasis in the original):

> The common thread will be the *Low-dimensional Topology*, a science that studies knots and manifolds of dimensions below five. The use of this name, Low-dimensional topology, began in the late seventies, but as a field it started long before. It was cultivated by figures as important as **Euler**, **Gauß**, **Riemann**, **Klein** and **Poincaré**, and by many others in the current century. But it was not until the year 1976 when this new science experimented an exceptional development, if not a complete revolution. A North-American mathematician with training in differential geometry, **W. Thurston**, began to use geometric methods that supplied a new impulse to the field. It is my intention here today to talk about these new methods and ideas.

In the academic term 1996-1997, JMMA was invited by the RAC to deliver the inaugural address. It was read on the 16th of October, 1996, and published as the RAC memoir [10]. The occasion was challenging, for it was expected that the speech should be accessible to a wide audience. The solution he adopted (to talk about the Poincaré problem) was to use a humorous cartoon-like verbal encoding just after testifying that

> it is a good lesson, that every scientist must learn, of how to communicate his knowledge in plain and understandable language for all.

As pointed at by the title, this memoir is an interesting essay on (some aspects of) the unity of mathematics, on the counterpoint between the finite and the infinite, between the discrete and the continuous. I am glad to say that Jaume Aguadé managed to capture faithfully the memoir style and content in his translation into Catalan that was published in 1997.

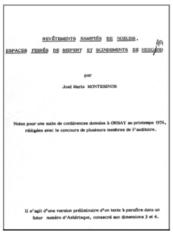
There are three more works published in the context of activities organized by the RAC aimed at more or less general publics. Two are about the history of geometry ([65] and [73]) and one about geometric crystallography [86]. They are included in the papers section because, as is the rule with JMMA, the reader will always find new poits of view, new ideas, new questions and often also new results.

Finally there are three recent papers published in RACSAM that seem to signal a new trend in his endeavor to contribute to the RAC mission: [103], [119] (joint with MTLI), and [121]. These papers are also special because they are focussed on new and promissing research lines.

Concluding sketches

The memoir [6] won the Prize of the RAC, but is was never published, apparently because it was too long (125 pages). Retrospectively, it seems clear that it was a pity. Taking as a starting point [139], the paper that had inspired his doctoral thesis, in the preface we read (our emphasis) that it:

> soon originated important researches on the topology of coverings that are branched along a link in S^3, and the aim of this memoir is to present *the part of these researches that is purely geometrical and in which we have obtained results or set forth some conjectures.* As a method for the description of results we adopt the different representations of the closed orientable three-dimensional manifolds, namely Heegaard's representation, the Dehn–Lickorish representation by means of surgery, and Alexander's representation using branched coverings. In our view we thus follow a geometrically logical sequence, which does not coincide with the historical unfolding. We omit proofs that involve an excessive algebraic machinery, and which otherwise can be found in the references at the end of the memoir. Instead, *some of the results are presented in a completely new fashion.* Finally, we illustrate the proofs, which are constructive for the most part, with *examples that are published for the first time.*


REVÊTEMENTS RAMIFIÉS DE NOEUDS,
ESPACES FIBRÉS DE SEIFERT ET SCINDEMENTS DE HEEGARD

par

José Maria MONTESINOS

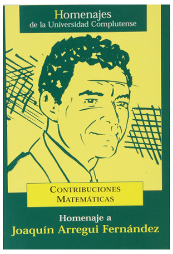
Notes pour une suite de conférences données à ORSAY au printemps 1976, rédigées avec le concours de plusieurs membres de l'auditoire.

Il s'agit d'une version préliminaire d'un texte à paraître dans un futur numéro d'Astérisque, consacré aux dimensions 3 et 4.

The memoir [7] was written by participants in the Spring 1976 Siebenmann's seminar at Orsay. JMMA was the speaker and the main focus of his lectures was, by the invitation agreement, the paper [15]. The papers [23] and [26] were also important. Michèle Audin and Francis Bohahon wrote a first version of the notes under the supervision of JMMA and in the year 1978 Siebenmann supervised a second version prepared by Michel Boileau and Yves Lemaire. Anyway, at the foot of the cover page we are told that it is a preliminary version of a text to be included in a future issue, devoted to dimensions 3 and 4, of the Astérisque collection (launched in 1973) of the French Mathematical Society. Regrettably, this publication did no happen and so this work has been, for all practical purposes, out of circulation ever since. It is the birth of Montesinos' links, a class that has remarkable ties with the Seifert fibrations, and, as indicated on page 31 (a note added by Siebenmann), "the main theorem [in the first chapter] has given rise to many other theorems that we omit for lack of time, and several among them have been discovered after these lectures".

As many others in his generation, JMMA's very early works were written in Spanish: The thesis [1], the papers [11]-[16], and the memoir [6] considered a moment ago. Then, starting with [17], he followed the trend of switching to English for the greatest part of his publications, particularly in the case of papers (96 out of 111), but with a number of important exceptions that can be tallid as follows: the books [2] and [4], which were born from visits to Mexico and Colombia, respectively; the memoirs other then [7] (this one in French, for the reasons already explained), which were produced in the context of RAC activities; the articles, whose general aim is to popularize some aspect of geometrical topology and its applications (but note that [125] appeared as a translation into Catalan and that [129] is a personal appreciation of Fico); the paper [33], included in the proceedings of the VII Spanish-Portuguese Mathematics Meeting; the Lecce (Italy) lecture [61]; the papers [65], [72] and [73], that are contributions to RAC projects ("History of mathematics in the XIX century", for the first two, and "Cultural horizons: the frontiers of science", for the third); and three more papers that we consider in next paragraph.

When the occasion has arisen to honor a teacher or a colleague, which usually happens in connection with some special birthday, as a rule JMMA has contributed with what he thinks is the only appropriate answer: a research paper. According to the references, the first two such occasions were the seventieth birthday of Luis Vigil Vázquez (1914-2003) and Francisco Botella Raduán (1915-1987), celebrated in 1984 and 1985, respectively. For the Vigil fest, he proposed a revision of the methods used to "'recognize S^3 from a Heegaard diagram" and an analysis of why they fail "for manifolds different from S^3". This paper was translated to English and published as [57]. In the case of Botella, he presented a different proof of a result of Sakuma. This paper can also be read in the original Spanish (included in a volume edited by E. Outerelo) and in English [53]. Next came the paper [62], this time in Spanish, dedicated to Antonio Plans Sanz de Bremond (1919-1997). It is a summary of a lecture JMMA imparted in 1986 at the UCM whose aim was to discuss the state of the art about conjectures relating the Heegaard genus of a 3-fold and different notions of rank. This was in part based on his paper [51].



The next three in this group are dedicated to three of his former teachers at the UCM and Faculty colleagues since 1986: [72], to José Javier Etayo Miqueo (1926-2012), in Spanish; [82], to Joaquín Arregui Fernández (1930-2012), in English; and [93], to Enrique Outerelo Domínguez (b. 1934), in Spanish. In the case of Etayo, he presented "the visual model of hyperbolic geometry", one of the aspects of his inquiries into that geometry, clearly related to activities in the RAC at that time, like [73]. The summary of the Arregui paper, reproduced by the MR, is a very clear description of his contributions (our emphasis): "Let M be a closed orientable 3-manifold. A *Dehn sphere* S is a 2-sphere immersed in M with only double curve and triple point singularities. S *fills* M if S defines a cell decomposition of M. *It is proven that every closed orientable* 3-*manifold has a filling Dehn sphere. Examples are given, and Johansson diagrams are proposed as a method for representing all closed orientable 3-manifolds*". The Outerelo paper got an excellet review (by D. Matei), but the English summary is perhaps the shortest description of its contents (our emphasis): "The fundaments of Fox's branched covering theory are *freshly exposed using inverse limits. Sufficient conditions for a branched covering to be onto, or open, or discrete are given. An example of a branched covering over* S^3 *with a nondiscrete fiber is constructed*". In the final part of his review, Matei adds: "A few intriguing open problems end the article: Are there non-surjective or non-open ramified coverings? Is there a ramified covering with a fiber homeomorphic to the Cantor set?".

More recently, in this honoring trend, it is the turn of colleagues of similar age or younger. The first in this series has been [112], dedicated to Juan Tarrés, and it will certainly continue with other names. H. M. Hilden published a nice review of the Tarrés paper, but again the summary is concise and to the point (our emphasis): "Under the framework of Fox spreads and its completions, a theory that generalizes coverings (*folding covering theory*) and a theory that generalizes branched coverings (*branched folding theory*) is defined and some properties are proved. *Two applications to* 3*-manifolds theory are given. A problem is stated*".

In addition to all the dedications we have mentioned, there are a number of other papers, sometimes joint papers, that are dedicated to a distinguished mathematician, again usually on the occasion around a special birthday. It is to note that this did not happen until JMMA had well established his international reputation in his specialty. The first was [44], dedicated "To Professor Arthur Stone" (Stone celebrated his seventieth birthday in 1986). Among the score of others, we mention here those professors that seem to occupy a more prominent place in JMMA's constellation: Deane Montgomery [52]; Joan S. Birman [58]; Tatsuo Homma, in his fifth Jupiterian year [63]; Yukio Matsumoto on his 60th birthday [88]; Laurent Siebenmann [90]; María Teresa Lozano Imízcoz, after 27 years of fruitful collaboration [102]; José Manuel Rodríguez Sanjurjo, in his 60th birthday [113]; Jozef Przytycki, in his 60th birthday [123]. There is also one that has a more personal meaning: "en recuerdo de mi padre, Lorenzo Montesinos" [87].

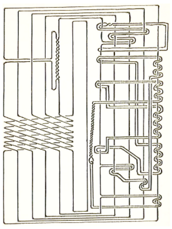
We have seen that a fraction of JMMA's works have been published in Spanish, even after switching to English as the main disseminating language. As it should be clear by what we have said so far, this has more of a personal option than of a constraint, for such works would no doubt have been accepted by good English journals. A similar comment can be advanced about the journals where his papers were sent. In fact, it is again easy to perceive in his production a laudable determination to support a variety of journals by sending them papers that would have no problem in being published in main stream English journals. Here are the main examples concerning journals published in Spain, Mexico and Colombia: BSMM [15, 94, 97, 104, 117, 118]; RACSAM [103, 119, 121]; RMHA [12, 13, 14]; RMUCM [87, 88, 90]; *Revista Colombiana de Matemáticas* [99, 102, 111]; *Collectanea Mathematica* [16, 40] and *Revista Academia Colombiana Ciencias* [96].

The ties of JMMA with Mexico and Colombia, not unlike those of many other country fellows for these and other Latinamerican nations, involve a deep emotional realization that he has expressed in words as follows ([4], Preface):

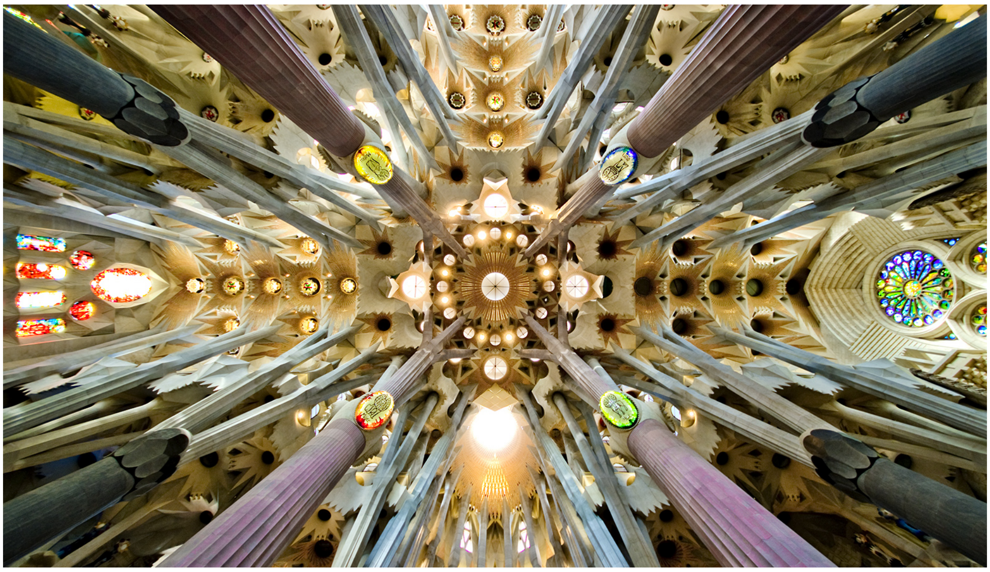
> The outcome and the development of the lectures were literally terrific: I was left open-mouthed on realizing the enthusiasm of the participants. There is in these dear people from Colombia and Iberoamerica something that we no longer have in the Old World and that is renewed each time one travels there. This 'something', very hard to explain, is a living reflection of its transparent skies and of the beautiful colors of the hummingbirds and flowers; of wonderful fruits.

> The friendship, endearing and warm; the faith, simple and strong, engine of hopes and forgiveness; the illusion for new worlds. It is enough to read their poets, if one is not there, or to converse with the people if you are lucky to be among them.

The papers and books of JMMA include many hundreds of pictures, usually drawn by his own hand. We have seen this in the case of the books [4], but we could take any other work, as for example the book [3], which has more than two hundred illustrations (counting numbered figures and graphics in tables). This profusion of graphics seems to be true, but perhaps to a lesser degree, for all low-dimensional topologists. In any case, such pictures form an integral part of how JMMA unfolds his thoughts and arguments. The drawings, and the mathematics behind them (particularly the subtle relations between the knots and the corresponding manifolds), point to an remarkably developed visualizing ability, an inner (mathematical, topological, geometrical) eye that seems utterly impossible to imagine by those lacking it.



On looking at such drawings, like the 'universal knot' in the image (Figure 23 in [48]), JMMA, and in general the low-dimensional topologists, somehow remind me of Antoni Gaudí. He was also endowed with a powerful visionary inner eye and his Sagrada Familia is to a large extent a geometrical monument. Thus there is some mathematical kindship with La Alhambra, the monument to which JMMA has devoted beautiful pages [3, 8, 126, 127, 130].



When people were expressing their astonishment at Gaudí's creativity, materialized in a panoply of forms and designs not yet organized in a whole, he always advised them to be patient, and then stated that "there will come folks from around the world to see them". Works such as those of JMMA, past, present and future, also require patience from the rest of us, but I am sure that there will come a time when mathematicians will feel the need, as this writer does, to learn more about it.

References

In the works authored by JMMA, a pointer to the MR is included whenever the corresponding review exists and is signed. With more that 110 reviews, their printout (including references) has over 75 pages.

Books

1972

[1] J. M. Montesinos: *Sobre la conjetura de Poincaré y los recubridores ramificados sobre un nudo*, Publicaciones de la Facultad de Ciencias, volumen A-165 (UCM, 1972). Tesis doctoral (1971). Cf. [141].

1982

[2] J. M. Montesinos: *Variedades de mosaicos.* VI ELAM. Instituto Politécnico Nacional de México, 1982.

1987

[3] J. M. Montesinos: *Classical Tessellations and Three-Manifolds.* Springer-Verlag, 1987. Translated to Japaneese, 1992. MR0915761 89d:57016 (R. Fenn).

2003

[4] J. M. Montesinos: *Calidoscopios y* 3*-Variedades.* Universidad Nacional de Colombia, 2003. Editado por D. M. Tejada y M. M. Toro.

2011

[5] Y. Matsumoto, J. M. Montesinos: *Pseudo-periodic Maps and Degeneration of Riemann Surfaces.* Springer, 2011. MR2839459 2012h:57039 (A. Papadopoulos).

Memoirs

1975

[6] J. M. Montesinos: *Sobre la representación de variedades tridimensionales.* Premio de la RAC, 1975 (no publicado).

1976

[7] J. M. Montesinos: *Revêtements ramifiés de noeuds. Espaces fibrés de Seifert et scindements de Heegaard, 1976.* Mimeographed notes of a one-month course delivered by the author at the Université Paris-Sud (Orsay).

1987

[8] J. M. Montesinos: *Caleidoscopios en la Alhambra.* Memorias de la RAC, Serie de ciencias exactas, 23 (1987). MR0916416 89i:01013 (J. S. Joel).

1990

[9] J. M. Montesinos: *Nudos, cristales y números: aspectos de la topología de baja dimensión.* Discurso leído en el acto de su recepción como académico numerario y contestación de José J. Etayo Miqueo. RAC, 1990.

1996

[10] J. M. Montesinos: *Números, combinatoria y nudos: de lo discreto a lo contínuo.* Discurso inaugural del Año Académico 1996-1997. RAC, 1996. Traducido al catalán por Jaume Aguadé: *Nombres, combinatòria i nusos: del discret al continu.* Butlletí de la Societat Catalana de Matemàtiques, 12(2) (1997) 7–45. MR1488295 99c:00001 (J. S. Joel).

Papers

1970

[11] J. M. Montesinos: *Una nueva clasificación de automorfismos en un espacio vectorial.* Gaceta Matemática 22(1) 22–35 (1970). MR0335542 49 323 (D. Carlson).

1972

[12] J. M. Montesinos: *Reducción de la conjetura de Poincaré a otras conjeturas geométricas.* RMHA 32(1-2) 33-51, 1972. MR0314038 47 2590 (H. B. Griffiths).

[13] J. M. Montesinos: *Una nota a un teorema de Alexander.* RMHA 32(4) 167–187, 1972. MR0385838 52 6697 (H. E. Debrunner).

1973

[14] J. M. Montesinos: *Una familia infinita de nudos representados no separables.* RMHA 33(4) 32–35, 1973. MR0385839 52 6698 (H. E. Debrunner).

[15] J. M. Montesinos: *Variedades de Seifert que son recubridores cíclicos ramificados de dos hojas.* BSMM 18(2) 1–32, 1973. MR0341467 49 6218 (H. E. Debrunner).

1974

[16] J. M. Montesinos: *Representaciones de enlaces en relación con recubridores dobles ramificados.* Collectanea Mathematica 25(2) 145–157, 1974. MR0420595 54 8609 (M. Gutierrez).

[17] J. M. Montesinos: *A representation of closed orientable 3-manifolds as 3-fold branched covering of S^3.* BAMS 80(5) 845–846, 1974. MR0358784 50 11243 (W. Whitten).

1975

[18] J. M. Montesinos: *3-Variétés qui ne sont pas des revêtements cycliques ramifiés sur S^3.* PAMS 47, 495–500, 1975. MR0353293 50 5777 (W. Whitten).

[19] J. M. Montesinos: *Corrigendum to* [18]. PAMS 53(2) 541, 1975. MR0383413 52 4294 (W. Whitten).

[20] J. M. Montesinos: *Surgery on links and double branched covers of S^3.* In [142] 227–259. Princeton University Press, 1975. MR0380802 52 1699 (R. J. Daverman).

1976

[21] J. M. Montesinos: *Three-manifolds as 3-fold branched covers of S^3.* QJM 27(1) 85–94, 1976. MR0394630 52 15431 (J. S. Birman).

[22] J. M. Montesinos: *Minimal plat representations of prime knots and links are not unique.* CJM, 28(1) 161–167, 1976. MR0410724 53 14470 (R. J. Daverman).

[23] J. S. Birman, F. González-Acuña, J. M. Montesinos: *Heegaard splittings of prime 3-manifolds are not unique.* Michigan Mathematical Journal, 23(2) 97–103, 1976. MR0431175 55 4177 (D. W. L. Sumners).

[24] H. Hilden, J. M. Montesinos, T. Thickstun: *Closed oriented 3-manifolds as 3-fold branched coverings of S^3 of special type.* PJM 65(1) 65–76, 1976. MR0420622 54 8635 (L. Neuwirth).

1977

[25] J. M. Montesinos: *Erratum: "Minimal plat representations of prime knots and links are not unique".* CJM 29(1) 224, 1977.

1978

[26] J. M. Montesinos: *4-manifolds, 3-fold covering spaces and ribbons.* TAMS 245, 453–467, 1978. MR0511423 80k:57001 (R. C. Kirby).

[27] F. González-Acuña, J. M. Montesinos: *Ends of knot groups.* AM (2) 108(1) 91–96, 1978. MR0488021 58 7598 (L. Neuwirth).

[28] H. Hilden, J. M. Montesinos: *A method of constructing 3-manifolds and its application to the computation of the μ-invariant.* In [143] Part 2, 61–69, 1978. MR0520523 80e:57003 (W. D. Neumann).

[29] Y. Matsumoto, J. M. Montesinos: *A note on contractible 3-manifolds* (Japanese). In *Proceedings of Kyoto Symposium,* Research Institute for Mathematical Sciences, Kyoto University, 1978.

1979

[30] J. M. Montesinos: *Heegaard diagrams for closed 4-manifolds.* In *Geometric Topology,* 219–237. Academic Press, 1979. MR0537732 80k:57022 (J. S. Birman).

1980

[31] H. M. Hilden, J. M. Montesinos: *Lifting surgeries to branched covering spaces.* TAMS 259(1) 157–165, 1980. MR0561830 81d:57002 (A. Edmonds).

[32] J. M. Montesinos: *A note on 3-fold branched coverings of S^3.* MPCPhS 88(2):321–325, 1980. MR0578276 81k:57002 (W. H. Heil).

[33] J. M. Montesinos: *Algunos aspectos de la teoría de cubiertas ramificadas.* Publicacions de la Secció de Matemàtiques de la Universitat Autònoma de Barcelona 21, 109–126, 1980 (Actas de la VII Reunión Hispano-Lusa de Matemáticas, Sant Feliu de Guíxols, Girona, España, 1980).

[34] J. S. Birman, J. M. Montesinos: *On minimal Heegaard splittings.* Michigan Mathematical Journal 27, 47–57, 1980. MR0555836 81b:57007 (H. M. Hilden).

[35] F. González-Acuña, J. M. Montesinos: *Non amphicheiral codimension two knots.* CJM 32, 185–194, 1980. MR0559794 81b:57017 (C. Kearton).

1982

[36] .F. González-Acuña, J. M. Montesinos: *Embedding knots in trivial knots.* BLMS 14, 238–240, 1982. MR0656605 83f:57014 (L. H. Kauffman).

1983

[37] J. M. Montesinos: *On twins in the four-sphere. I.* QJM 34(134) 171–179, 1983. MR0698205 86i:57025a (L. H. Kauffman).

[38] J. M. Montesinos: *Representing 3-manifolds by a universal branching set.* MPCPhS 94(1) 109–123, 1983. MR0704805 85c:57001 (J. H. Rubinstein).

[39] F. González-Acuña, J. M. Montesinos: *Quasiaspherical knots with infinitely many ends.* Commentarii Mathematici Helvetici 58(1) 257–263, 1983. MR0705536 85c:57023 (J. G. Ratcliffe).

[40] H. Hilden, M. T. Lozano, J. M. Montesinos: *The Whitehead link, the Borromean rings and the knot* 9_{46} *are universal.* Collectanea Mathematica 34(1) 19–28, 1983. MR0747855 86b:57001 (G. Burde).

[41] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *All three-manifolds are pullbacks of a branched covering* S^3 *to* S^3*.* TAMS 279(2) 729–735, 1983. MR0709580 85b:57015 (J. S. Birman).

[42] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *Universal knots.* BAMS 8(3) 449–450, 1983. MR0693959 84g:57002 (W. D. Neumann).

1984

[43] J. M. Montesinos: *On twins in the four-sphere. II.* QJM 35(137) 73–83, 1984. MR0734666 86i:57025b (L. H. Kauffman).

1985

[44] J. M. Montesinos: *Lectures on 3-fold simple coverings and 3-manifolds.* Contemporary Mathematics 44, 157–177, 1985. Proceedings of the 1982 Rochester Conference on "Combinatorial methods in Topology and Algebraic Geometry". MR0813111 87b:57002 (G. Burde).

[45] J. M. Montesinos: *A note on moves and on irregular coverings of* S^4*.* Contemporary Mathematics 44, 345–349, 1985. Proceedings of 1982 Rochester Conference on "Combinatorial methods in Topology and Algebraic Geometry". MR0813123 87h:57006 (U. Hirsch).

[46] J. M. Montesinos: *Lectures on branched coverings.* Atti del Convegno Studio Sulla Geometria delle Varietà Differenziabili 127–167, 1985 (edited by I. Cataneo Gasparini).

[47] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On knots that are universal.* Topology 24(4) 499–504, 1985. MR0816529 87a:57010 (W. H. Heil).

[48] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *Universal knots.* In *Knot theory and manifolds (Vancouver, 1983),* LNM 1144 (Springer, 1985), 25–59. MR0823281 87g:57008 (W. H. Heil).

1986

[49] C. McA. Gordon, J. M. Montesinos: *Fibred knots and disks with clasps.* Mathematische Annalen 275(3) 405–408, 1986. MR0858286 88e:57007 (J. Hempel).

[50] J. M. Montesinos: *A note on twist spun knots.* PAMS, 180–184, 1986. MR0848900 87h:57023 (C. Gordon).

[51] J. M. Montesinos: *Note on a result of Boileau-Zieschang.* Lectures Notes Series of the LMS, 112, 241–252, 1986 ("Low-dimensional topology and Kleinian groups", Coventry/Durham, 1984). MR0903868 89f:57002 (N. V. Ivanov).

[52] J. M. Montesinos, W. Whitten: *Constructions of two-fold branched covering spaces.* PJM 125(2) 415–446, 1986.

1987

[53] J. M. Montesinos: *On 3-manifolds having surface bundles as branched coverings.* PAMS 101(3) 555–558, 1987. English versión of a paper published in *Contribuciones matemáticas en honor de* Francisco Botella Raduán (editado por Enrique Outerelo, UCM 1986) 197–201. MR0873520 88b:57004 (S. H. Weintraub).

[54] M. Boileau, F. González-Acuña, J. M. Montesinos: *Surgery on double knots and symmetries.* Mathematische Annalen 276(2) 323–340, 1987. MR0870971 88h:57003 (J. H. Przytycki).

[55] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *Non-simple universal knots.* MPCPhS 102(01) 87–95, 1987. MR0886438 88j:57006 (S. Suzuki).

[56] H. M. Hilden, M. T. Lozano, J. M. Montesinos, W. C. Whitten: *On universal groups and three-manifolds.* Inventiones mathematicae 87(3) 441–456, 1987. MR0874030 88e:57016 (V. G. Turaev).

1988

[57] J. M. Montesinos: *Two questions on Heegaard diagrams of S^3.* PAMS 102(2) 421–425, 1988. English version of a paper published in *Contribuciones matemáticas en honor de* Luis Vigil Vázquez (Universidad de Zaragoza 1984) 233–239. MR0921010 89f:57012 (C. Gordon).

[58] J. M. Montesinos, C. Safont: *On the Birman invariants of Heegaard splittings.* PJM 132(1) 113–142, 1988. MR0929585 89d:57017 (H. Hendriks).

[59] V. Chumillas, J. M. Montesinos: *The homology of cyclic and irregular dihedral coverings branched over homology spheres.* Mathematische Annalen 280(3) 483–500, 1988. MR0936324 89i:57001 (Y. Shinohara).

[60] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On the universal group of the Borromean rings.* In *Differential Topology,* LNM 1350 (Proceedings of the 1987 Siegen conference on Differential Topology, Springer, 1988), 1–13. MR0979330 89m:57002 (W. Whitten).

1989

[61] J. M. Montesinos: *La discrepancia entre el rango y el genero de Heegaard de una 3-variedad.* Note di Matematica IX-Supplemento, 101–117, 1989. Lecture delivered at the 1989 Lecce Conference on "Differential Geometry and Topology". MR1154135 93e:57030 (J. Vrabec).

1990

[62] J. M. Montesinos: *Género de Heegaard y rango de una 3-variedad.* In *Contribuciones matemáticas en homenaje a* Antonio Plans Sanz de Bremond (Universidad de Zaragoza 1990) 191–199.

1991

[63] Y. Matsumoto, J. M. Montesinos: *A proof of Thurston's uniformization theorem of geometric orbifolds.* Tokyo Journal of Mathematics 14(1) 181–196, 1991. MR1108165 92g:57005 (W. Dunbar).

[64] J. M. Montesinos, H. R. Morton: *Fibred links from closed braids.* PLMS (3) 62(1) 167–201, 1991. MR1078219 91m:57003 (A. Edmonds).

1992

[65] J. M. Montesinos: *Geometrías no-euclídeas: Gauss, Lobachevskii y Bolyai.* In *Historia de las matemáticas en el siglo XIX, Parte I* (RAC, 1992), 65–114. MR1469501 98g:01043 (D. Ştefănescu).

[66] H. M. Hilden, M.T. Lozano, J. M. Montesinos: *On the Borromean orbifolds: geometry and arithmetic.* In *Topology '90* (Columbus, OH, 1990). Publications of the Ohio State University Mathematical Research Institute, vol. 1, 133–167, 1992. MR1184408 93h:57021 (B. N. Apanasov).

[67] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *The arithmeticity of the figure eight knot orbifolds.* In *Topology '90* (Columbus, OH, 1990). Publications of the Ohio State University Mathematical Research Institute, vol.-1, 169–183, 1992. MR1184409 93k:57012 (B. N. Apanasov).

[68] H. M. Hilden, M.T. Lozano, J. M. Montesinos: *A characterization of arithmetic subgroups of* $SL(2,\mathbb{R})$ *and* $SL(2,\mathbb{C})$. Mathematische Nachrichten 159(1) 245–270, 1992. MR1237113 94i:20088 (A. W. Reid).

[69] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On the character variety of group representations of a 2-bridge link* $p/3$ *into* $PSL(2,\mathbb{C})$. BSMM (2) 37(1-2) 241–262, 1992. MR1317576 96a:57018 (R. Fenn).

1993

[70] F. González-Acuña, J. M. Montesinos: *On the character variety of group representations in* $SL(2,\mathbb{C})$ *and* $PSL(2,\mathbb{C})$. Mathematische Zeitschrift 214(1) 627–652, 1993. MR1248117 94k:57022 (D. McCullough).

[71] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *Universal 2-bridge knot and link orbifolds.* JKT 2(2) 141–148, 1993. MR1227007 94i:57012 (F. Paulin).

1994

[72] J. M. Montesinos: *El modelo visual de la geometría hiperbólica.* In *Contribuciones Matemáticas en honor de* José Javier Etayo Miqueo (UCM 1994) 467–482.

[73] J. M. Montesinos: *La cuestión de la consistencia de la geometría hiperbólica.* In *Historia de las matemáticas en el siglo XIX, Parte II* (RAC, 1994), 213–232.

[74] Y. Matsumoto, J. M. Montesinos: *Pseudo-periodic homeomorphisms and degeneration of Riemann surfaces.* BAMS 30(1) 70–75, 1994. MR1217354 94h:30057 (Y. Imayoshi).

1995

[75] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On the arithmetic 2-bridge knots and link orbifolds and a new knot invariant.* JKT 4(1) 81–114, 1995. MR1321291 96a:57019 (A. W. Reid).

[76] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On a remarkable polyhedron geometrizing the figure eight knot cone manifolds.* Journal of Mathematical Sciences University of Tokyo 2(3) 501–561, 1995. MR1382519 96m:57021 (W. Dunbar).

1996

[77] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On volumes and Chern-Simons invariants of geometric 3-manifolds.* Journal Mathematical Sciences University Tokyo 3(3) 723–744, 1996. MR1432115 98h:57056 (D. C. Calegari).

[78] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *Volumes and Chern-Simons invariants of cyclic coverings over rational knots.* In *Topology and Teichmüller spaces* (Katinkulta, 1995), 31–35, 1996. MR1659808 2000a:57022 (Y. Rong).

1997

[79] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *The arithmeticity of certain torus bundle cone 3-manifolds and hyperbolic surface bundle 3-manifolds; and an enhanced arithmeticity test.* In *Tokyo KNOTS '96* (World Scientific, 1997), 73–80. MR1664951 2000a:57029 (D. McCullough).

[80] M. T. Lozano, J. M. Montesinos: *Geodesic flows on hyperbolic orbifolds, and universal orbifolds.* PJM 177(1) 109–147, 1997. MR1444776 98c:58125 (L. Mosher).

1999

[81] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *The Chern–Simons invariants of hyperbolic manifolds via covering spaces.* BLMS 31(3) 354–366, 1999. MR1673415 2000c:57022 (D. Auckly).

2000

[82] J. M. Montesinos: *Representing 3-manifolds by Dehn spheres.* In *Contribuciones matemáticas en homenaje a* Joaquín Arregui Fernández (UCM 2000) 239–247.

[83] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On the character variety of tunnel number 1 knots.* JLMS (2) 62(3) 938–950, 2000. MR1794296 2001i:57009 (Y. Q. Wu).

[84] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On the character variety of periodic knots and links.* MPCPhS 129(3) 477–490, 2000. MR1780499 2001g:57012 (N. Chbili).

2001

[85] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *The arithmetic structure of a universal group.* Atti del Seminario Matematico e Fisico dell'Università di Modena 49(supplemento) 1–14, 2001. MR1881088 2003f:20069 (K. N. Jones).

2002

[86] J. M. Montesinos: *La cristalografía geométrica.* In *Horizontes Culturales: las fronteras de la Ciencia* (RAC, 2002), 97–112.

[87] J. M. Montesinos: *Representing open 3-manifolds as 3-fold branched coverings.* RMUCM 15(2) 533–542, 2002. MR1951824 2004a:57026 (A. Edmonds).

2003

[88] J. M. Montesinos: *Uncountably many wild knots whose cyclic branched covering are* S^3. RMUCM 16(1) 329–344, 2003. MR2031886 2004m:57027 (A. Edmonds).

[89] J. M. Montesinos: *Embedding strings in the unknot.* Journal Mathematical Sciences University of Tokyo 10, 631–660, 2003.

[90] J. M. Montesinos: *Open 3-manifolds, wild subsets of S^3 and branched coverings.* RMUCM 16(2) 577–600, 2003. MR2032934 2005a:57011 (A. Edmonds).

[91] J. M. Montesinos: *On Cantor sets in 3-manifolds and branched coverings.* QJM 54(2) 209–212, 2003. MR1989872 2004d:57003 (D. Lima Gonçalves).

[92] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *Character varieties and peripheral polynomials of a class of knots.* JKT 12(08) 1093–1130, 2003. MR2017984 2004i:57005 (D. Auckly).

2004

[93] J. M. Montesinos: *Propiedades topológicas básicas de las cubiertas ramificadas de Fox.* In *Contribuciones Matemáticas en Homenaje del Profesor* Enrique Outerelo Domínguez (UCM 2004), 315–332. MR2212974 2007b:57004 (D. Matei).

[94] H. L. Hilden, M. T. Lozano, J. M. Montesinos: *On 2-universal knots.* BSMM (3) 10 (Special Issue) 239–253, 2004. MR2199351 2006i:57012 (M. Kranjc).

[95] H. M. Hilden, J. M. Montesinos, D. M. Tejada, M. M. Toro: *Knots, butterflies and 3-manifolds.* Disertaciones del Seminario de Matemáticas Fundamentales de la UNED 33, 1–22, 2004.

[96] H. M. Hilden, J. M. Montesinos, D. M. Tejada, M. M. Toro: *Mariposas y 3-variedades.* Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales 28, 71–78, 2004. MR2065841 2005d:57007 (S. Francaviglia).

2005

[97] J. M. Montesinos: *Branched coverings after Fox.* BSMM (3) 11(1) 19–64, 2005. MR2198590 2006i:57005 (A. Edmonds).

[98] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *Peripheral polynomials of hyperbolic knots.* Topology and its Applications 150(1) 267–288, 2005.

[99] H. M. Hilden, J. M. Montesinos, D. Tejada, M. Toro: *Representing 3-manifolds by triangulations of S^3: a constructive approach.* Revista Colombiana de Matemáticas 39(2) 63–86, 2005. MR2218370 2007c:57029 (F. Spaggiari).

2006

[100] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On hyperbolic 3-manifolds with an infinite number of fibrations over S^1.* MPCPhS 140(1) 79–93, 2006. MR2197577 2006i:57035 (J. D. Masters)

[101] H. M. Hilden, J. M. Montesinos, D. M. Tejada, M. M. Toro: *Fox coloured knots and triangulations of S^3.* MPCPhS 141(03) 443–463, 2006. MR2281408 2007h:57004 (A. Papadopoulos).

2007

[102] J. M. Montesinos: *Open 3-manifolds and branched coverings: a quick exposition.* Revista Colombiana de Matemáticas 41(2) 287–302, 2007. MR2599123 2011a:57006 (A. Edmonds).

2008

[103] J. M. Montesinos: *Around the Borromean link.* RACSAM 102(1) 75–87, 2008.

[104] G. Brumfiel, H. M. Hilden, M. T. Lozano, J. M. Montesinos, E. Ramírez-Losada, H. Short, D. Tejada, M. Toro: *Three manifolds as geometric branched coverings*

of the three sphere. BSMM (3) 14(2) 263–282, 2008. arXiv:0710.1960. MR2667329 2011h:57004 (W. H. Heil).

[105] G. Brumfiel, H. M. Hilden, M. T. Lozano, J. M. Montesinos, E. Ramírez-Losada, H. Short, D. Tejada, M. Toro: *On finite index subgroups of a universal group.* BSMM (3) 14(2) 283-301, 2008. arXiv:0710.5835. MR2667330 2011h:57005 (W. H. Heil).

2010

[106] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On universal hyperbolic orbifold structures in S^3 with the Borromean rings as singularity.* Hiroshima Mathematical Journal 40(3) 357–370, 2010. MR2766667 2012d:57042 (T. Li).

[107] J. M. Montesinos: *The knot engulfing property of 3-manifolds.* Atti del Seminario Matematico e Fisico dell'Università di Modena e Reggio Emilia 57, 135–140, 2010. MR2894979 2012m:57040 (W. H. Heil).

2011

[108] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On representations of 2-bridge knot groups in quaternion algebras.* JKT 20(10) 1419–1483, 2011. arXiv:1001.3546v1.

[109] J. J. Monzón, A. G. Barriuso, L. L. Sánchez-Soto, J. M. Montesinos: *Geometrical interpretation of optical absorption.* Physical Review A 84(2) 023830, 2011.

[110] H. M. Hilden, J. M. Montesinos, J. M. Tejada, D. M. Toro: *Artifacts for stamping symmetric designs.* American Mathematical Monthly 118(4) 327–343, 2011.

2012

[111] H. M. Hilden, J. M. Montesinos, D. Tejada, M. M. Toro: *On the classification of 3–bridge links.* Revista Colombiana de Matemáticas 46(2) 113–144, 2012. MR3007934 (P. Lopes).

[112] J. M. Montesinos: *Branched folded coverings and 3-manifolds.* In *Contribuciones matemáticas en honor de* Juan Tarrés (UCM 2012) 295–315. MR3060537 (H. M. Hilden).

2013

[113] J. M. Montesinos: *On integral quadratic forms having commensurable groups of automorphisms.* Hiroshima Mathematical Journal 43(3) 371–411, 2013. MR3161323 (A. Sładek).

[114] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On representations of 2-bridge knot groups in quaternion algebras, II.* JKT 22(1) 125–140, 2013.

[115] J. J. Monzón, A. G. Barriuso, J. M. Montesinos, L. L. Sánchez-Soto: *Geometrical aspects of $\mathcal{PT}$-invariant transfer matrices.* Physical Review A 87(1) 012111 (4 pages), 2013.

2014

[116] J. M. Montesinos: *Addendum to* [113]. Hiroshima Mathematical Journal 44, 341–350, 2014.

[117] J. M. Montesinos: *On binary integral quadratic forms having isometric groups of automorphisms.* BSMM 20(2) 199–236, 2014.

[118] J. M. Montesinos: *On some open 3-manifolds that are branched coverings of the Poincaré homology sphere.* BSMM 20(2) 505–521, 2014.

[119] M. T. Lozano, J. M. Montesinos: *On the degeneration of some 3-manifold geometries via unit groups of quaternion algebras.* RACSAM 109:669-715, 2015.

[120] M. T. Lozano, J. M. Montesinos: *Geometric cone manifold structures on $T_{p/q}$, the result of p/q surgery in the left handed trefoil knot T.* To appear in JKT (accepted 24 August 2015).

2015

[121] J. M. Montesinos: *On odd rank integral quadratic forms: canonical representatives of projective classes and explicit construction of integral classes with square-free determinant.* RACSAM 109(1) 199–245, 2015.

[122] J. M. Montesinos: *Erratum to: "On odd rank integral quadratic forms: canonical representatives of projective classes and explicit construction of integral classes with square-free determinant".* RACSAM 109:759-760, 2015.

[123] J. M. Montesinos: *On some groups related with Fox–Artin wild arcs.* Preprint, 2015.

[124] J. M. Montesinos: *On arithmetic line geometry: the automorphs a quadratic line complex. Projective classes of quaternary integral quadratic forms of index 2.* Preprint, 2015.

Articles

1984

[125] J. M. Montesinos: *Punts de vista sobre el problema de Poincaré.* El desenvolupament de les Matemàtiques en el segle XIX (edited by Manuel Castellet, Institut d'Estudis Catalans, Arxius de la Secció de Ciències, 1984), 23–40.

1987

[126] J. M. Montesinos: *Caleidoscopios en la Alhambra.* Boletín de la Sociedad Castellana Puig Adam de Profesores de Matemáticas 13, 29–45, 1987.

[127] J. M. Montesinos: *Caleidoscopios y grupos cristalográficos en La Alhambra* Epsilón, 1987, v. [145] 9–36.

2001

[128] J. M. Montesinos: *Los anillos de Borromeo.* In [151] (Academia de Ciencias de Zaragoza, 2001) 69–73.

2002

[129] J. M. Montesinos: Francisco González Acuña*: Semblanza Humana e Intelectual.* Carta informativa de la Sociedad Matemática Mexicana, Septiembre 2002, 1–2. Reproduced, with an English translation, in BSMM (3) 10 (Special Issue):xiii–xix, 2004.

2008

[130] J. M. Montesinos: *Geometría y Topología en la Alhambra.* In *Aspectos matemáticos en la ciencia y en la sociedad* (edición de P. Jiménez, Instituto de España, 2008), 111–141.

2009

[131] J. M. Montesinos: *Cristalografía y arte.* RACSAM 103(1) 207–222, 2009.

2010

[132] J. M. Montesinos: *Grupos cristalográficos y topología en Escher.* RACSAM 104(1) 27–47, 2010. MR2894911 2012k:20094 (C. Corrales-Rodrigáñez).

2012

[133] J. M. Montesinos: *Un problema bimilenario: la geometría no euclidiana hiperbólica.* RACSAM 105(2) 217–226, 2013.

2013

[134] J. M. Montesinos: *En torno a los números perfectos.* Boletín de la Sociedad Puig Adam, 94, 12–28, Junio 2013. Número especial dedicado al Profesor José Javier Etayo Miqueo.

External references

[135] H. Seifert and W. Threlfall: *Lehrbuch der Topologie.* Teubner, 1934. Translation into English (by M. A. Goldman) published in 1980 by Academic Press together with the translation by Wolfgam Heil of H. Seifert: *Topology of 3-dimensional fibered spaces.* The joint text was edited by Joan S. Birman and Julian Eisner.

[136] M. K. Fort, Jr., editor: *Topology of 3-manifolds and related topics.* Prentice Hall, 1962 (Dover edition 2010, with an introduction by D. S. Silver).

[137] R. H. Fox: *A quick trip through knot theory.* In [136], 120–167.

[138] R. H. Fox: *Some problems in knot theory.* In [136], 168–170.

[139] R. H. Fox: *Construction of simply connected 3-manifolds.* In [136], 213–217.

[140] R. H. Crowell and R. H. Fox: *Introduction to Knot Theory.* Ginn/Boston, 1963. In 1977 it was reprinted by Springer-Verlag, with corrections and a preface by R. H. Crowell. In 2008 it was reprinted by Dover.

[141] R. H. Fox: *A note on branched cyclic coverings of spheres.* RMHA 32, 158–166, 1972. Contains a condensed version of Montesinos' PhD thesis [1].

[142] L. P. Neuwirth, editor: *Knots, Groups and 3-Manifolds*, Annals of Mathematics Studies vol. 84, Princeton University Press / University of Tokio Press, 1975.

[143] R. J. Milgram, editor: *Algebraic and geometric topology* (Stanford University, August 2-21, 1976). Proceedings of Symposia in Pure Mathematics XXXII (2 parts), American Mathematical Society, 1978.

[144] G. Burde and H. Zieschang: *Knots.* Studies in Mathematics 5. Walter de Gruyter, 1985.

[145] Asociación de Profesores de Matemáticas de Andalucía, editor: *La Alhambra.* εpsilon (volumen especial). Consejería de Cultura de la Junta de Andalucía, 1987.

[146] L. H. Kauffman: *On Knots.* Annals of Mathematical Studies 115. Princeton University Press, 1987.

[147] E. B. Vinberg, editor: *Geometry, II.* Encyclopedia of Mathematical Sciences 29, Springer, 1988.

[148] J. G. Ratcliffe: *Foundations of hyperbolic manifolds.* GTM 149. Springer, 1994.

[149] S. Xambó: Sobre puntos y coordenadas. *In* Contribuciones Matemáticas en honor de José Javier Etayo Miqueo (UCM 1994) 61–93.

[150] W. Thurston: *Three-Dimensional Geometry and Topology.* Mathematical series 35. Princeton University Press, 1997.

[151] M. T. Lozano, editor: *Matemáticas en el año 2000.* Monografías de la Academia de Ciencias de Zaragoza v.19, 2001.

[152] Departamento de Geometría y Topología, editor: *Contribuciones matemáticas en honor de* José María Montesinos Amilibia (UCM 2009).

[153] S. Xambó: *José María Montesinos Amilibia.* In *"El Árbol de las Matemáticas: Galería RSME-Universia de Matemáticas, Ciencia y Tecnología".* (`http://www.arbolmat.com/`).

[154] M. T. Lozano. *Algunas contribuciones matemáticas del Profesor* José María Montesinos Amilibia. In *A mathematical tribute to Professor José María Montesinos Amilibia* (UCM 2015).

Acronyms

AM	Annals of Mathematics
AMS	American Mathematical Society
BAMS	Bulletin of the American Mathematical Society
BLMS	Bulletin of the London Mathematical Society
BSMM	Boletín de la Sociedad Matemática Mexicana
CJM	Canadian Journal of Mathematics
ELAM	Escuela Latinoamericana de Matemáticas
GTM	Graduate Texts in Mathematics
IAS	Institute for Advances Study
JKT	Journal of Knot Theory and its Ramifications
JLMS	Journal of the London Mathematical Society
JMMA	José María Montesinos Amilibia
LMS	London Mathematical Society
LNiM	Lecture Notes in Mathematics
MPCPhS	Mathematical Proceedings of the Cambridge Philosophical Society
MSRI	Mathematical Sciences Research Institute
MTLI	María Teresa Lozano Imízcoz
PAMS	Proceedings of the American Mathematical Society
PJM	Pacific Journal of Mathematics
PLMS	Proceedings of the London Mathematical Society
RAC	Real Academia de Ciencias Exactas, Físicas y Naturales de Madrid
RACSAM	Revista de la RAC, Serie A, Matematicas
QJM	Quarterly Journal of Mathematics
RMHA	Revista Matemática Hispano-Americana
RMUCM	Revista Matemática Complutense
TAMS	Transactions of the American Mathematical Society
UAM	Universidad Autónoma de Madrid
UB	Universitat de Barcelona
UCM	Universidad Complutense de Madrid
UNED	Universidad Nacional de Educación a Distancia
UZ	Universidad de Zaragoza

Proof of the rigidity of model nilpotent Lie algebras by means of the Internal Set Theory

J. M. Ancochea Bermúdez and R. Campoamor Stursberg*

Departmento de Geometría y Topología
Universidad Complutense de Madrid
28040 Madrid, Spain
ancochea@ucm.es rutwig@ucm.es

Dedicado al profesor D. José María Montesinos, amigo y maestro, con motivo de su jubilación académica.

ABSTRACT

In this note, we provide a proof based on the Internal Set Theory of Nelson concerning the existence of solvable Lie algebras having a nilradical with arbitrary characteristic sequence.

2010 Mathematics Subject Classification: 17B30, 17B05.
Key words: geometric rigidity, Lie algebras, nilpotency, Internal Set Theory.

1. Introduction

The interest on (finite dimensional) rigid Lie algebras emerges naturally from the study of the geometry of the variety $\mathcal{L}^n$ of Lie algebras of dimension n, and more specifically, for determining the irreducible components [5, 10]. Whereas the rigidity of semisimple algebras follows easily from their structure, the rigidity problem for non-semisimple Lie algebras is much more involved, as an algebra can be rigid in two ways: geometrically or algebraically, the latter implying the former by the inverse being generally false [14, 15]. As complete classifications of solvable Lie algebras are non-existent, the study of rigidity must be carried out either considering special types of Lie algebras or by means of indirect arguments, such as the theory of weights [1, 4, 7, 8, 15].

*Author partially supported by the research project MTM2013-43820-P of the MINECO.

The objective of this note is to prove, using techniques from Non-standard Analysis, that for any characteristic sequence there exists at least a nilpotent Lie algebra $\mathfrak{n}$ such that it appears as the nilradical (maximal nilpotent ideal) of a solvable rigid Lie algebra $\mathfrak{r}$.

Unless otherwise stated, any Lie algebra considered in this work is defined over $\mathbb{C}$.

1.1. Generalities

Given a Lie algebra $\mathfrak{g}$, we denote the Lie algebra formed by the derivations of $\mathfrak{g}$ by $Der(\mathfrak{g})$.

Definition 1.1 *Let $\mathfrak{g}$ be a real Lie algebra of dimension n. An external torus of derivations is any Abelian subalgebra of $Der(\mathfrak{g})$ the generators of which are semisimple.*

It is called maximal torus, if it is maximal for the inclusion relation. Maximal tori have the same dimension that we call the rank of $\mathfrak{g}$ and denote by $\mathrm{r}(\mathfrak{g})$. The Mal'cev theorem ensures that maximal tori are pairwise conjugated [12]. With this result, the generic structure of solvable Lie algebras has been studied in detail (see e.g. [3, 6, 9]).

Let $\mathcal{L}^n$ denote the class of n-dimensional Lie algebras $\mathfrak{g} = \left(\mathbb{C}^n, [\,,\,]_{\mathfrak{g}}\right)$. The the general linear group $GL(n,\mathbb{C})$ acts naturally on $\mathcal{L}^n$ by changes of basis:

$$(f.\mathfrak{g})(X,Y) = f^{-1}\left([f(X), f(Y)]_{\mathfrak{g}}\right),\ f \in GL(n,\mathbb{C}),\ X, Y \in \mathfrak{g}.. \tag{1.1}$$

Using the tensor formalism, we denote $[X,Y]_{\mathfrak{g}} = \mu(X,Y)$ for all $X, Y \in \mathfrak{g}$, where μ is a skew-symmetric tensor of type $(2,1)$ satisfying the Jacobi identity. The orbit $\mathcal{O}(\mathfrak{g})$ of $\mathfrak{g}$ is formed by those Lie algebras that are isomorphic to $\mathfrak{g}$.

Definition 1.2 *A Lie algebra $\mathfrak{g}$ is rigid if the orbit $\mathcal{O}(\mathfrak{g})$ is an open set of $\mathcal{L}^n$ with respect to the Euclidean topology.*

In this sense, the notion of rigidity constitutes an important tool in the analysis of the irreducible components of the variety $\mathcal{L}^n$ of n-dimensional Lie algebras [4, 10]. One of the difficulties in the analysis of rigidity is to find algebraic criteria that ensure the topological property above, having in mind that these normally constitute sufficient but not necessary conditions [15]. An important result in this direction is given by Nijenhuis and Richardson [14]:

Proposition 1 *If the second Chevalley cohomology group $H^2(\mathfrak{g},\mathfrak{g})$ vanishes, then $\mathfrak{g}$ is a rigid Lie algebra.*

1.2. The Carles decomposition theorem

A notion central to the structure of complex rigid Lie algebras is their decomposability, as shown by R. Carles in [6].

Definition 1.3 *A Lie algebra $\mathfrak{g}$ is decomposable if it admits the decomposition*

$$\mathfrak{g} = \mathfrak{s} \overrightarrow{\oplus} \left(\mathfrak{t} \overrightarrow{\oplus} \mathfrak{n} \right), \tag{1.2}$$

where $\mathfrak{s}$ is a Levi subalgebra, $\mathfrak{n}$ the nilradical and $\mathfrak{t}$ a subalgebra formed by $ad_{\mathfrak{g}}$-semisimple elements and such that $[\mathfrak{s} + \mathfrak{t}, \mathfrak{t}] = 0$ holds. The preceding decomposition is called the normal decomposition of $\mathfrak{g}$.

Theorem 1.1 *Any rigid Lie algebra $\mathfrak{g}$ admits a normal decomposition. Moreover, $\mathfrak{g}$ satisfies one of the following conditions:*

1. *The radical $\mathfrak{r}$ of $\mathfrak{g}$ is not nilpotent and $\dim Der(\mathfrak{g}) = \dim \mathfrak{g}$. Moreover, if $\operatorname{codim}_{\mathfrak{g}} [\mathfrak{g}, \mathfrak{g}] > 1$, then $\mathfrak{g}$ has no center and any derivation is inner.*
2. *The radical $\mathfrak{r}$ is nilpotent and one of the following subcases holds:*
   - *(a) $\mathfrak{g} = [\mathfrak{g}, \mathfrak{g}]$, i.e., $\mathfrak{g}$ is a perfect Lie algebra.*
   - *(b) $\mathfrak{g} = \mathfrak{g}' \oplus \mathbb{C}$ is the direct sum of a perfect Lie algebra $\mathfrak{g}'$ with only inner derivations and $\mathbb{C}$.*
   - *(c) $\mathfrak{g} \neq [\mathfrak{g}, \mathfrak{g}]$ and possesses no direct non-zero Abelian factor, then $\mathfrak{t}_e(\mathfrak{g}) = \mathfrak{t}_e(\mathrm{J}) = 0$ for any codimension 1 ideal J.*

2. The rank theorem

In this section we recall a method applicable to solvable Lie algebras that provides a sufficiency criterion to check rigidity [3]. The procedure is independent on cohomology, hence valid for the geometric and algebraic rigidity of Lie algebras.

Let $\mathfrak{g}$ be a solvable decomposable Lie algebra. We fix a maximal external torus $\mathfrak{t}$. Let X be a non-zero vector such that $ad_{\mathfrak{g}}X$ belongs to $\mathfrak{t}$.

Definition 2.1 *X is called a regular element if the dimension of*

$$V_0(X) = \{Y \mid [X, Y] = 0\} \tag{2.1}$$

is minimal.

Let X be a regular element and let $p = \dim V_0(X)$. Consider a basis of eigenvectors $\{X, Y_1, \ldots, Y_{n-p}, X_1, \ldots, X_{p-1}\}$ for the adjoint operator $\mathrm{ad}_{\mathfrak{g}}(X)$ such that $\{X, X_1, \ldots, X_{p-1}\}$ is a basis of $V_0(X)$, $\{Y_1, \ldots, Y_{n-p}, X_1, \ldots X_{k_0}\}$ is a basis of the nilradical $\mathfrak{n}$ of $\mathfrak{g}$ and $\{X_{k_0+1}, \ldots, X_{p-1}\}$ are such that $\mathrm{ad}_{\mathfrak{g}}(X_j) \in \mathfrak{t}$ for $j = k_0 + 1, \ldots, p$.

We denote by $\mathrm{I}(X)$ be the sequence of eigenvalues of the operator $ad_{\mathfrak{g}}X$ with respect to $(Y_1,\ldots,Y_{n-p},X_1,\ldots,X_{p-1})$. It follows that $\mathrm{I}(X)=\{\lambda_1,\ldots,\lambda_{n-p},0,\ldots,0\}$ with $[X,Y_i]=\lambda_i Y_i$.
If $p\neq n$, the vector X does not belong to the derived ideal of $\mathfrak{g}$, and the vectors $[X_i,X_j]$, $[X_i,X_j]$ and $[Y_i,Y_j]$ don't contain X. If $p=n$, $\mathfrak{g}$ is easily seen to be nilpotent [3].

Suppose that $\mathfrak{g}$ is solvable non-nilpotent. The root system of $\mathfrak{g}$ associated to the basis $(X,Y_1,\ldots,Y_{n-p},X_1,\ldots,X_{p-1})$ is the linear system $\mathbf{S}$ defined by the following equations:

$$\begin{array}{llll} x_i+x_j=x_k & \text{if the } X_k\text{-coordinate of} & [X_i,X_j] & \text{is non-zero,}\\ y_i+y_j=y_k & \text{if the } Y_k\text{-coordinate of} & [Y_i,Y_j] & \text{is non-zero,}\\ x_i+y_j=y_k & \text{if the } Y_k\text{-coordinate of} & [X_i,Y_j] & \text{is non-zero,}\\ y_i+y_j=x_k & \text{if the } X_k\text{-coordinate of} & [Y_i,Y_j] & \text{is non-zero.} \end{array} \tag{2.2}$$

With these notations, the rank theorem proved in [3] states the following:

Theorem 2.1 *If* $\operatorname{rank}(\mathbf{S})\neq\dim\mathfrak{n}-1$*, then the* $\mathfrak{g}$ *cannot be a rigid Lie algebra.*

An important property used in the proof is that the rank of a root system for $\mathfrak{g}$ is independent on the basis of eigenvectors considered, hence it constitutes an invariant of the isomorphism class of $\mathfrak{g}$. Two additional relevant properties concerning the rank theorem are enumerated:

1. If $\mathfrak{g}$ is rigid, then there always exists a regular element X such that the adjoint operator $ad_{\mathfrak{g}}X$ is diagonal with integer eigenvalues.
2. If the $\mathfrak{g}=\mathfrak{t}\overrightarrow{\oplus}\mathfrak{n}$ is rigid, then $\mathfrak{t}$ is a maximal external torus of $\mathfrak{g}$.

The last property is a consequence of the Mal'cev theorem [12]. In particular, the dimension of $\mathfrak{t}$ satisfies the upper bound

$$\dim\mathfrak{n}-\dim[\mathfrak{n},\mathfrak{n}]\geq\dim\mathfrak{t}. \tag{2.3}$$

3. The model nilpotent Lie algebra $\mathfrak{n}_c$

The characteristic sequence $c(\mathfrak{n})$ of a nilpotent Lie algebra, originally conceived as an additional invariant to separate isomorphism classes within the variety of nilpotent Lie algebra laws, turns out to be of great practical use in rigidity theory [8], as it allows a systematical analysis of solvable rigid algebras according to the generic structure of their nilradical.
Given a nilpotent Lie algebra $\mathfrak{n}$, for a non-zero element $X\in\mathfrak{n}\setminus[\mathfrak{n},\mathfrak{n}]$ we consider the decreasing sequence of dimensions of the Jordan blocks of the linear operator $\mathrm{ad}(X)$.

$$c(X)=(c_1(X),c_2(X),\ldots,c_k(X),1),\quad c_i(X)\geq c_{i+1}(X)\geq 1. \tag{3.1}$$

Definition 3.1 *The characteristic sequence of the nilpotent Lie algebra* $\mathfrak{n}$ *is defined as*

$$c(\mathfrak{n}) = \sup \{c(X) \mid X \in \mathfrak{n} \setminus [\mathfrak{n}, \mathfrak{n}]\} \tag{3.2}$$

It follows that for an arbitrary nilpotent Lie algebra possessing the characteristic sequence $c(\mathfrak{n}) = (n_1, n_2, \ldots, n_k, 1)$, we can always find a basis adapted to it $\left\{X_1, \ldots, X_{n_1+1}, \ldots, X_{n_1+n_2+1}, \ldots, X_{n_1+\cdots+n_{k-1}+1}, \ldots, X_{n_1+\cdots+n_k+1}\right\}$ and satisfying the following commutators

$$\begin{aligned}
[X_1, X_j] &= & X_{j+1},\ 2 \leq j \leq n_1; \\
[X_1, X_{n_1+j}] &= & X_{n_1+1+j},\ 2 \leq j \leq n_2; \\
\vdots & & \vdots \\
\left[X_1, X_{n_1+\cdots+n_{k-2}+j}\right] &= & X_{n_1+\cdots+n_{k-2}+1+j},\ 2 \leq j \leq n_{k-1}; \\
\left[X_1, X_{n_1+\cdots+n_{k-1}+j}\right] &= & X_{n_1+\cdots+n_{k-1}+1+j},\ 2 \leq j \leq n_k.
\end{aligned} \tag{3.3}$$

The remaining brackets $[X_i, X_j]$ for $2 \leq i, j$ are related by means of the Jacobi condition

$$[X_1, [X_i, X_j]] + [X_j, [X_1, X_i]] + [X_i, [X_j, X_1]] = 0. \tag{3.4}$$

For the special case $c(\mathfrak{n}) = (n-1, 1)$, the Lie algebra $\mathfrak{n}$ is called the filiform model algebra and denoted by L_n.

We observe that if $n_j = 1$ for some j, then $\mathfrak{n}_c$ splits into a direct sum of Lie algebras. In order to handle only with indecomposable nilpotent algebras, in the following we require, unless otherwise stated, that $n_j > 1$ holds for any $j = 1, \ldots, k$.

Definition 3.2 *Let* $k > 1$*. The nilpotent Lie algebra* $\mathfrak{n}_c$ *with commutators (3.3) and* $[X_i, X_j] = 0$ *for* $i, j \geq 2$ *is called the model nilpotent Lie algebra for the characteristic sequence* $c(\mathfrak{n}) = (n_1, n_2, \ldots, n_k, 1)$*.*

Using the rank theorem, it is straightforward to verify that, given a characteristic sequence $c(\mathfrak{n}) = (n_1, n_2, \ldots, n_k, 1)$, the rank of $\mathfrak{n}_c$ equals $\operatorname{rank}(\mathfrak{n}_c) = k+1$.

We now prove that for any possible characteristic sequence characteristic sequence $c(\mathfrak{n}) = (n_1, n_2, \ldots, n_k, 1)$, there exists a solvable rigid Lie algebra possessing $\mathfrak{n}_c$ as nilradical. To this extent, we use the perturbation theory inferred from the Internal Set Theory.[1]

Proposition 2 *For any characteristic sequence* $c(\mathfrak{n}) = (n_1, n_2, \ldots, n_k, 1)$*,* $\mathfrak{n}_c$ *is isomorphic to the nilradical of a solvable rigid Lie algebra* $\mathfrak{r}_c$ *of rank* $k+1$*.*

[1]See the Appendix for the elementary definitions and axioms. Further detail can be found in [11, 13].

Proof. Let μ_0 be the structure tensor underlying the nilpotent Lie algebra $\mathfrak{n}_c$. The root system **S** associated to μ_0 is

$$\begin{array}{rcll} \lambda_1 + \lambda_j &=& \lambda_{j+1}, & 2 \leq j \leq n_1 \\ \lambda_1 + \lambda_{n_1+j} &=& \lambda_{n_1+j+1}, & 2 \leq j \leq n_2 \\ \vdots & & & \vdots \\ \lambda_1 + \lambda_{n_1+\cdots+n_{k-1}+j} &=& \lambda_{n_1+\cdots+n_{k-1}+j+1}, & 2 \leq j \leq n_k \end{array} \tag{3.5}$$

A basis of solutions to this system is given by the vectors

$$\begin{array}{l} v_1 = (1,2,\ldots,n_1,n_1+1,n_2,\ldots,n_1+\cdots+n_k+1) \\ v_2 = \left(0,1,\ldots^{(n_1)}\ldots,1,0,\ldots 0,\ldots,0,\ldots 0,\ldots,0\right) \\ v_3 = \left(0,0,\ldots,0,1,\ldots^{(n_2)}\ldots,1,0,\ldots 0,\ldots,0,\ldots 0\right) \\ \vdots \\ v_k = \left(0,0,\ldots,0,0,\ldots,0,1,\ldots^{(n_{k-1})}\ldots,1,0,\ldots 0\right) \\ v_{k+1} = \left(0,0,\ldots,0,0,\ldots,0,0,\ldots,0,1,\ldots^{(n_k)}\ldots,1\right) \end{array} \tag{3.6}$$

Now let μ be a perturbation of μ_0. The linear operator $ad_\mu T_1$ has eigenvalues

$$\{0,\lambda_1,\lambda_2,\ldots,\lambda_n,\nu_1,\ldots,\nu_k\}, \tag{3.7}$$

where $n = \sum_{l=1}^k n_l + 1$ and the eigenvalues are such that $\lambda_i \sim i$ for $1 \leq i \leq n$ and $\nu_j \sim 0$ for $1 \leq j \leq k$. Since for any standard element $Y \in \mathfrak{t} = \langle T_1, T_2, \ldots, T_{k+1}\rangle$ there exists an index $i \in \{1,\ldots,n\}$ such that $\mu_0(Y,X_i) = aX_i$ with $a \in \mathbb{R}^*$ standard, we have that $\nu_1 = \cdots = \nu_k = 0$. Moreover, since $\lambda_i \neq \lambda_j$ for $i \neq j$, the operator $ad_\mu(T_1)$ is diagonalizable. Since any of the eigenspaces of eigenvalue σ of $ad_\mu(T_1)$ have as shadow the eigenspace of $ad_{\mu_0}(T_1)$ corresponding to the shadow of the eigenvalue σ (see e.g. [11]), we can always find a basis of eigenvectors

$$B = \{Y_1,\ldots,Y_n,T_1,U_2,\ldots,U_{k+1}\} \tag{3.8}$$

of $ad_\mu(X)$ such that

$$\begin{array}{ll} {}^oY_i & = X_i,\ 1 \leq i \leq n \\ {}^oU_j & = T_j,\ 2 \leq j \leq k+1 \end{array}$$

On the other hand, since $\mu(Y_i,Y_j)$ is an eigenvector associated to the eigenvalue $\lambda_i + \lambda_j$, then $\lambda_i + \lambda_j = \lambda_{i+j}$ holds if $\mu(Y_i,Y_j) \neq 0$. Applying the Jacobi identity we successively deduce that

$$\mu(U_i,U_j) = 0,\ 2 \leq i,j \leq k+1 \tag{3.9}$$

The eigenvalues of the linear operator $ad_\mu(T_1+U_i)$ are pairwise distinct, from which we further obtain that $ad_\mu(U_i)$ is a diagonal operator over the basis (3.8). We therefore have

$$\begin{array}{rcll} \mu(Y_i,Y_j) &=& \lambda_{i+j}Y_{i+j}, & 1 \leq i,j \leq n \\ \mu(T_1,Y_i) &=& \lambda_i Y_i, & 1 \leq i \leq n \\ \mu(U_j,Y_i) &=& a_{ji}Y_i, & 1 \leq i \leq n,\ j \geq 2 \end{array} \tag{3.10}$$

where $\lambda_{1+i} \sim 1$ for all $i \neq \{1, n_1+1, n_2+1, \dots, n_{k-1}+1\}$ and $a_{ji} \sim 1$ para $j = 2, \dots, k+1$ and $i = n_1 + \cdots + n_{j-2} + 2, \dots, n_1 + \cdots + n_{j-2} + n_{j-1} + 1$ and $a_{ji} \sim 0$ otherwise.

We conclude that the vectors

$$\begin{array}{l} w_1 = (\lambda_1, \lambda_2, \dots, \lambda_n, 0, \dots, 0) \\ w_2 = (a_{21}, a_{22}, \dots, a_{2n}, 0, \dots, 0) \\ \vdots \\ w_{k+1} = (a_{k+1,1}, a_{k+1,2}, \dots, a_{k+1,n}, 0, \dots, 0) \end{array} \tag{3.11}$$

constitute a basis of the root system $\mathbf{S}$ associated to the Lie algebra μ_0. This enables us to find a basis $\{V_1, \dots, V_{k+1}\}$ of $\langle T_1, U_2, \dots, U_{k+1}\rangle$ such that

$$\begin{array}{rl} \mu(V_1, Y_i) = i\, Y_i, & 2 \leq i \leq n \\ \mu\left(V_j, Y_{f(j)+l}\right) = Y_{f(j)+l}, & 2 \leq j \leq k+1,\ 2 \leq l \leq n_{j-1} \end{array} \tag{3.12}$$

where $f(j) = n_1 + n_2 + \cdots + n_{j-2} + 1$. The Jacobi identities imply the identities

$$X_{1+j} = \mu_0(X_1, X_j) \sim \mu(Y_1, Y_j) = \lambda_{1+j} Y_{1+j}, \tag{3.13}$$

where $j = 2, \dots, n_1, n_1+2, \dots, n_1+n_2, n_1+n_2+2, \dots, n_1+\cdots n_{k-1}+2, \dots, n_1+\cdots+n_k$, as well as

$$0 = \mu_0(X_i, X_j) \sim \mu(Y_i, Y_j) = \lambda_{i+j} Y_{1+j},\ 2 \leq i, j \leq n. \tag{3.14}$$

Finally, considering the change of basis

$$\begin{array}{c} Z_1 = Y_1,\ Z_2 = Y_2, \\ Z_{n_1+2} = Y_{n_1+2},\ Z_{n_1+3} = \lambda_{n_1+3} Y_{n_1+3}, \dots, Z_{n_1+n_2+1} = \prod_{l=3}^{n_2+1} \lambda_{n_1+l} Y_{n_1+n_2+1} \\ \vdots \\ Z_{n_1+\cdots+n_{k-1}+2} = Y_{n_1+2},\ Z_{n_1+\cdots+n_{k-1}+3} = \lambda_{n_1+\cdots+n_{k-1}+3} Y_{n_1+3}, \dots, \\ Z_n = \prod_{l=3}^{n_k} \lambda_{n_1+\cdots+n_{k-1}+l} Y_n, \end{array} \tag{3.15}$$

it follows at once that μ has the same structure constants as μ_0, showing that $\mu \simeq \mu_0$. □

It should be observed that actually, a stronger result can be proved, namely, that the model nilpotent Lie algebras arise as the nilradical of a solvable algebraically rigid Lie algebra $\mathfrak{r}_c$ of rank $k+1$. This sharpened version uses the Hochschild-Serre spectral sequence, as well as the rigidity criteria of Nijenhuis and Richardson. The proof will appear elsewhere.

4. Seven-dimensional model algebras

As an application of the result, we determine the solvable rigid Lie algebras the nilradical of which is a seven dimensional model nilpotent algebra. Eleven cases are given, some of which have already appeared in classifications in low dimension [1, 2, 9].

1. $\mathfrak{r}_{(6,1)} = \mathbb{C}\langle T_1, T_2\rangle \overrightarrow{\oplus} \mathfrak{n}_{(6,1)}$:

$$[X_1, X_2] = X_3, \quad [X_1, X_3] = X_4, \quad [X_1, X_4] = X_5, \quad [X_1, X_5] = X_6,$$
$$[X_1, X_6] = X_7, \quad [T_1, X_\alpha] = \alpha X_\alpha, \quad \alpha = 1, \dots, 7; \quad [T_2, X_\beta] = X_\beta,$$
$$\beta = 2, \dots, 7.$$

2. $\mathfrak{r}_{(5,1,1)} = \mathbb{C}\langle T_1, T_2, T_3\rangle \overrightarrow{\oplus} \mathfrak{n}_{(5,1,1)}$:

$$[X_1, X_2] = X_3, \quad [X_1, X_3] = X_4, \quad [X_1, X_4] = X_5, \quad [X_1, X_5] = X_6,$$
$$[T_1, X_\alpha] = \alpha X_\alpha, \quad \alpha = 1, \dots, 6; \quad [T_2, X_\beta] = X_\beta, \quad \beta = 2, \dots, 6;$$
$$[T_3, X_7] = X_7.$$

3. $\mathfrak{r}_{(4,2,1)} = \mathbb{C}\langle T_1, T_2, T_3\rangle \overrightarrow{\oplus} \mathfrak{n}_{(4,2,1)}$:

$$[X_1, X_2] = X_3, \quad [X_1, X_3] = X_4, \quad [X_1, X_4] = X_5, \quad [X_1, X_6] = X_7,$$
$$[T_1, X_\alpha] = \alpha X_\alpha, \quad \alpha = 1, \dots, 7; \quad [T_2, X_\beta] = X_\beta, \quad \beta = 2, \dots, 7,$$
$$[T_3, X_\gamma] = X_\gamma, \quad \gamma = 6, 7.$$

4. $\mathfrak{r}_{(4,1,1,1)} = \mathbb{C}\langle T_1, T_2, T_3, T_4\rangle \overrightarrow{\oplus} \mathfrak{n}_{(4,1,1,1)}$:

$$[X_1, X_2] = X_3, \quad [X_1, X_3] = X_4, \quad [X_1, X_4] = X_5, \quad [T_1, X_\alpha] = \alpha X_\alpha,$$
$$\alpha = 1, \dots, 5; \quad [T_2, X_\beta] = X_\beta, \quad \beta = 2, \dots, 5; \quad [T_3, X_6] = X_6,$$
$$[T_4, X_7] = X_7.$$

5. $\mathfrak{r}_{(3,3,1)} = \mathbb{C}\langle T_1, T_2, T_3\rangle \overrightarrow{\oplus} \mathfrak{n}_{(3,3,1)}$:

$$[X_1, X_2] = X_3, \quad [X_1, X_3] = X_4, \quad [X_1, X_5] = X_6, \quad [X_1, X_6] = X_7,$$
$$[T_1, X_\alpha] = \alpha X_\alpha, \quad \alpha = 1, \dots, 7; \quad [T_2, X_\beta] = X_\beta, \quad \beta = 2, 3, 4;$$
$$[T_3, X_\gamma] = X_\gamma, \quad \gamma = 5, 6, 7.$$

6. $\mathfrak{r}_{(3,2,1,1)} = \mathbb{C}\langle T_1, T_2, T_3, T_4\rangle \overrightarrow{\oplus} \mathfrak{n}_{(3,2,1,1)}$:

$$[X_1, X_2] = X_3, \quad [X_1, X_3] = X_4, \quad [X_1, X_5] = X_6, \quad [T_1, X_\alpha] = \alpha X_\alpha,$$
$$\alpha = 1, \dots, 7; \quad [T_2, X_\beta] = X_\beta, \quad \beta = 2, 3, 4; \quad [T_3, X_\gamma] = X_\gamma$$
$$\beta = 5, 6; \quad [T_4, X_7] = X_7.$$

7. $\mathfrak{r}_{(3,1,1,1,1)} = \mathbb{C}\langle T_1, T_2, T_3, T_4, T_5\rangle \overrightarrow{\oplus} \mathfrak{n}_{(3,1,1,1,1)}$:

$$[X_1, X_2] = X_3, \quad [X_1, X_3] = X_4, \quad [T_1, X_\alpha] = \alpha X_\alpha, \quad \alpha = 1, \dots, 7;$$
$$[T_2, X_\beta] = X_\beta, \quad \beta = 2, 3, 4; \quad [T_k, X_{k+2}] = X_{k+2}, \quad k = 3, 4, 5.$$

8. $\mathfrak{r}_{(2,2,2,1)} = \mathbb{C}\langle T_1, T_2, T_3, T_4\rangle \overrightarrow{\oplus} \mathfrak{n}_{(2,2,2,1)}$:

$$[X_1, X_2] = X_3, \quad [X_1, X_4] = X_5, \quad [X_1, X_6] = X_7, \quad [T_1, X_\alpha] = \alpha X_\alpha,$$
$$\alpha = 1, \dots, 7; \quad [T_2, X_\beta] = X_\beta, \quad \beta = 2, 3; \quad [T_3, X_\gamma] = X_\gamma,$$
$$\gamma = 4, 5; \quad [T_4, X_\delta] = X_\delta, \quad \delta = 6, 7.$$

9. $\mathfrak{r}_{(2,2,1,1,1)} = \mathbb{C}\langle T_1, T_2, T_3, T_4, T_5 \rangle \overrightarrow{\oplus} \mathfrak{n}_{(2,2,1,1,1)}$:

$$\begin{array}{llll} [X_1, X_2] = X_3, & [X_1, X_4] = X_5, & [T_1, X_\alpha] = \alpha X_\alpha, & \alpha = 1, \ldots, 7; \\ [T_2, X_\beta] = X_\beta, & \beta = 2, 3; & [T_3, X_\gamma] = X_\gamma, & \gamma = 4, 5; \\ [T_k, X_{2+k}] = X_{2+k}, & k = 1, 2. & & \end{array}$$

10. $\mathfrak{r}_{(2,1,1,1,1,1)} = \mathbb{C}\langle T_1, T_2, T_3, T_4, T_5, T_6 \rangle \overrightarrow{\oplus} \mathfrak{n}_{(2,1,1,1,1,1)}$:

$$\begin{array}{llll} [X_1, X_2] = X_3, & [T_1, X_\alpha] = \alpha X_\alpha, & \alpha = 1, \ldots, 7; & [T_2, X_\beta] = X_\beta, \\ \beta = 2, 3; & [T_k, X_{1+k}] = X_{1+k}, & k = 3, 4, 5, 6; & \end{array}$$

11. $\mathfrak{r}_{(1,1,1,1,1,1,1)} = \mathbb{C}\langle T_1, \ldots, T_7 \rangle \overrightarrow{\oplus} \mathfrak{n}_{(1,1,1,1,1,1,1)}$:

$$[T_k, X_k] = X_k, \quad k = 1, \ldots, 7.$$

Appendix. Internal Set Theory

The I.S.T. theory constitutes a conservative extension of the Zermelo-Fraenkel set theory with the axiom of choice (ZFC), to which a new predicate called "standard" (symbol: st) and three additional axioms are added. According to [13], we call a formula *internal* in case it does not involve the new predicate "standard"; otherwise we call it *external*. The three new axioms added to the ZFC theory are the following [13]:

Idealization principle (I):

$$\forall^{st} \text{finite} x' \exists y \forall x \in x' A \longleftrightarrow \exists y \forall^{st} x A. \tag{4.1}$$

Standardization Principle (S):

$$\forall^{st} X \, \exists^{st} Y \, \forall^{st} z \, [z \in Y \longleftrightarrow (z \in X) \wedge A] \tag{4.2}$$

Transfer principle (T): Let A be an internal formula whose only free variables are $x, t_1, \ldots, t_n$. Then

$$\forall^{st} t_1 \cdots \forall^{st} t_n \left[\forall^{st} x A \longleftrightarrow \forall x A \right] \tag{4.3}$$

The transfer principle states that if A is an internal formula and all parameter have standard values, if A holds for all standard x, then it holds for all x.

Using the predicate standard, we can make the following definitions, where the variables range over $\mathbb{R}$. For the complex case $\mathbb{C}$, the situation is analogous, taking into account the real and imaginary parts, respectively.

- x is called *infinitesimal* in case for all standard $\varepsilon > 0$ we have $|x| < \varepsilon$.

- x is called *limited* in case for some standard r we have $|x| \leq r$.
- $x \sim y$ [x and y are infinitely close] in case $x - y$ is infinitesimal.

In particular, the axiom **(S)** establishes that for any limited real x there is a unique standard ${}^0x$ such that $x \sim {}^0x$, called the *shadow* of x. The shadow of linear subspaces can be defined in analogous way.

The relevant results for perturbations of algebraic structures are the following [11]:

Proposition A1. Let n be standard and V be a linear subspace of $\mathbb{C}^n$. Then ${}^0V$ is a linear subspace of $\mathbb{C}^n$ such that $\dim V = \dim^0 V$.

Proposition A2. Let $P_0 = \lambda_i x^i \in \mathbb{C}_n[x]$ be a complex polynomial of degree n with standard complex coefficients λ_i and let $P = \mu_i x^i$ be a polynomial of degree n such that $\lambda_i \sim \mu_i$. If $\alpha_1, \ldots, \alpha_n$ are the roots of P and $\beta_1, \ldots, \beta_n$ the roots of P_0, then $\beta_j = {}^0\alpha_j$ for $1 \leq j \leq n$.

Proposition A3. Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be a standard linear map and T_0 a perturbation of T. Then the shadow ${}^0\lambda$ of an eigenvalue λ of T_0 is an eigenvalue of T. In addition, if v is a limited eigenvector of T_0 associated to the eigenvalue λ, then ${}^0v$ is an eigenvector of T with eigenvalue ${}^0\lambda$.

References

[1] J. M. Ancochea: *On the rigidity of solvable Lie algebras.* In *Deformation Theory of Algebras and Structures and Applications* (Il Ciocco, 1986), NATO Adv. Sci. Inst. Ser. C. Math. Phys. Sci. **247,** Kluwer Acad. Publ., 1988, 403–445.

[2] J. M. Ancochea, R. Campoamor-Stursberg: *On Lie algebras whose nilradical is (n-p)-filiform.* Comm. Algebra **29** (2001) 427-450.

[3] J. M. Ancochea, M. Goze: *Le rang du système linéaire des racines d'une algèbre de Lie rigide résoluble complexe* Comm. Algebra **20** (1992) 875–887.

[4] R. Carles: *Rigidité dans la variété des algèbres de Lie et composantes irréductibles.* C. R. Acad. Sci. Paris **286** (1978) 1223–1226.

[5] R. Carles: *Sur certaines classes d'orbites ouvertes dans les variétés d'algèbres de Lie.* C. R. Acad. Sci. Paris **293** (1981), 545–547.

[6] R. Carles: *Sur la structure des algèbres de Lie rigides.* Ann. Inst. Fourier **34** (1984) 65–82.

[7] G. Favre: *Système de poids sur une algèbre de Lie nilpotente.* Manuscripta Math. **9** (1973) 53–90.

[8] M. Goze, J. M. Ancochea: *Algèbres de Lie rigides.* Nederl. Akad. Wetensch. Indag. Math. **47** (1985) 397–415.

[9] M. Goze, J. M. Ancochea: *Algèbres de Lie rigides dont le nilradical est filforme* C. R. Acad. Sic. Paris **312** (1991) 21-24.

[10] A. A. Kirillov, Yu. A. Neretin: *The variety A_n of n-dimensional Lie algebra structures.* Amer. Mat. Soc. Transl. **137** (1987) 21–30.

[11] R. Lutz, M. Goze: *Nonstandard Analysis: a Practical Guide with Applications.* Spinger LNM **881,** 1981.

[12] A. I. Mal'cev: *Solvable Lie algebras.* Izv. Akad. Nauk SSSR **9** (1945) 329–356.

[13] E. Nelson: *Internal Set Theory: a new approach to Nonstandard Analysis.*Bull. Amer. Math. Soc. **83** (1977) 1165–1198.

[14] A. Nijenhuis, R. W. Richardson: *Cohomology and deformations in graded Lie algebras* Bull. Amer. Math. Soc. **72** (1966) 1–29.

[15] G. Rauch: *Effaçement et déformation.* Ann. Inst. Fourier **22** (1972) 239–269.

Una caracterización de las proyecciones de Lagrange

Jorge L. Andrés y Jesús Otero*

Facultad de Matemáticas
Universidad Complutense de Madrid
28040 Madrid, Spain
joryepe@gmail.com oteroj@ucm.es

Resumen

En esta nota presentamos una nueva caracterización de las proyecciones de Lagrange en su forma canónica, que establece la dependencia lineal entre la transformación conforme asociada $f(z)$ y su derivada pre-schwarziana, esta última relacionada con las curvaturas de las imágenes de los meridianos y de los paralelos terrestres.

2010 Mathematics Subject Classification: 30C35, 86A30.
Key words: Cartography, conformal mappings, pre-Schwarzian derivative.

1. Introducción y preliminares

La célebre memoria de Lagrange [6] *Sur la construction des cartes géographiques*, publicada en las Memorias de la Academia de Ciencias de Berlín de 1779, es pieza fundamental en el desarrollo de la cartografía matemática durante el siglo XIX, y así fue considerada hasta las primeras décadas del siglo XX. Sin embargo, las proyecciones conformes descubiertas por Lagrange pasan hoy prácticamente inadvertidas y están relegadas al grupo de las proyecciones misceláneas, tal y como sucede en el conocido albúm de proyecciones cartográficas de Snyder y Voxland [7]. Una de las causas de este "olvido" pudiera deberse a la presentación, en algunos textos, de las proyecciones de Lagrange mediante una *latitud modificada*, introducida genialmente por Tissot [8, Complément, § 111] con la intención de asemejar sus ecuaciones a las de la proyección estereográfica ecuatorial, expresadas estas últimas en función de la latitud y longitud geográficas.

*Autor apoyado por el proyecto MTM2014-57113-P

Estas proyecciones de Lagrange son las proyecciones conformes de la esfera (en general de una superficie de revolución, tal y como hizo el matemático francés) que representan los meridianos y los paralelos como arcos de circunferencias, excepto el meridiano y el paralelo de un punto z_0 (*centro de la proyección*), que se transforman en rectas (ortogonales). En la Figura 1 mostramos el aspecto típico de una proyección de Lagrange con centro en un punto del ecuador.

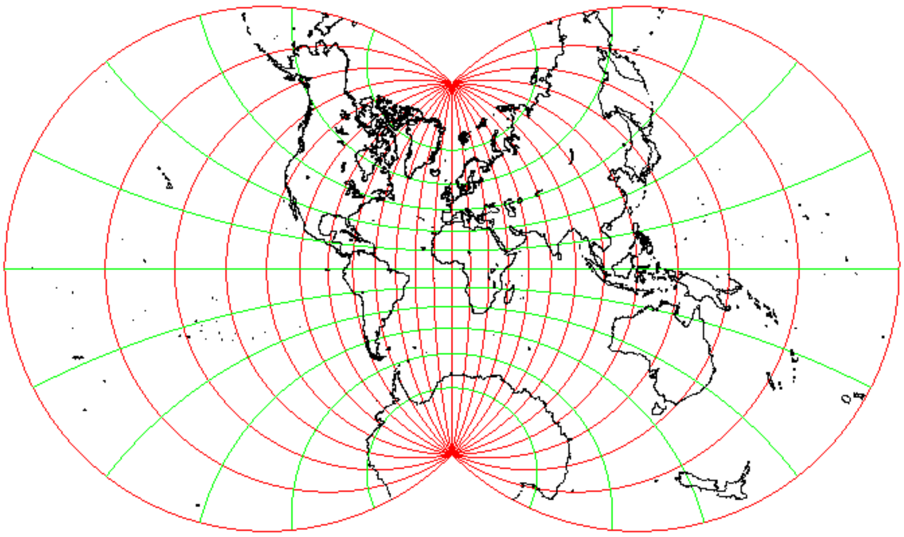


Figura 1: Proyección de Lagrange: en rojo las imágenes de los meridianos, y en verde las imágenes de los paralelos. El ángulo interior en los polos es 270º.

Siguiendo a Lagrange, en este artículo parametrizamos la esfera S^2 con la longitud geográfica $\lambda \in (-\pi, \pi]$ y la latitud isométrica $q \in (-\infty, \infty)$, definida a partir de la latitud geográfica φ por la identidad

$$q = \log \operatorname{tg} \left[\frac{\pi}{4} + \frac{\varphi}{2} \right] . \tag{1.1}$$

La función (1.1) es estrictamente creciente y por tanto biyectiva entre $(-\pi/2, \pi/2)$ y $\mathbb{R}$. En estas coordenadas isométricas $(\lambda\,, q)$ el elemento de arco de la esfera toma la forma

$$ds^2 = r^2(d\lambda^2 + dq^2)\,,$$

donde $r = \operatorname{sech} q$ es el radio del paralelo de latitud isométrica q, igual a $\cos\varphi$ en función de la latitud geográfica. El mapa que corresponde a esta parametrización es la carta de Mercator que se obtiene con la proyección $x \in S^2 \to z = \lambda + iq$. Al igual que en geometría diferencial utilizamos la notación (u, v) para designar a las

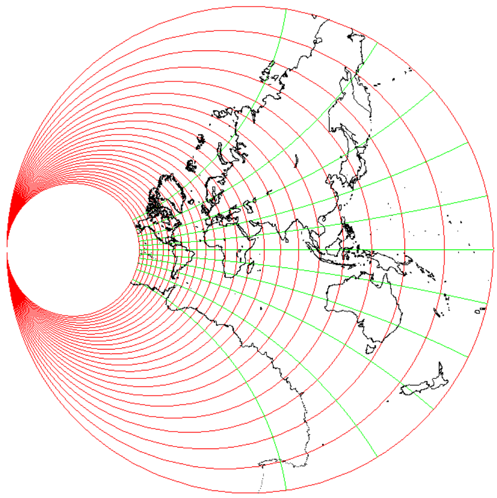


Figura 2: Imágenes de los meridianos (rojo) y paralelos (verde) en la transformación racional lineal $f(z) = (7\pi/4 - z)^{-1}$.

coordenadas (λ, q). A diferencia de lo habitual, en cartografía se reserva la notación $w \equiv x + iy$ para designar al punto proyectado $f(u + iv)$.

La teoría de proyecciones cartográficas conformes, de la esfera o del elipsoide de revolución terrestre, trata pues sobre ciertas tranformaciones conformes $f(z)$ de la banda infinita $B = (-\pi, \pi] \times \mathbb{R}$, donde interesa principalmente el análisis de la representación de la red cartesiana (cartesian net) por coincidir con la red de meridianos y paralelos terrestres. Otra cuestión básica en cartografía es tratar de minimizar la distorsión del mapa, es decir, la que introduce la composición de la proyección de Mercator (primer mapa) y la transformación $f(z)$. El problema de encontrar la "mejor" proyección conforme para un país fue abordado por P.L. Chebyshev, quien publicó sus resultados en 1856 en la Academia Imperial de las Ciencias de San Petersburgo, incluyendo un estudio particular de las proyecciones de Lagrange [3].

En el problema de Lagrange, las transformaciones buscadas son aquellas que representan la red cartesiana en una red de circunferencias (ortogonales), y que llamamos *transformaciones conformes de Lagrange*. Excluimos en este artículo las proyecciones con meridianos rectilíneos (circunferencias de radio infinito): proyecciones cilíndricas conformes (Mercator), proyección estereográfica polar y proyecciones cónicas conformes de Lambert. También excluimos las que se obtienen mediante transformaciones racionales lineales del plano z, $f(z) = (az + b)/(cz + d)$, ya que representan los polos norte y sur geográficos en el mismo punto a/c (Figura 2), mientras que aquí centramos nuestro interés en las proyecciones *bipolares*, es decir, aquellas en las cuales los polos tienen representación finita y distinta.

En la citada memoria [6, §10], Lagrange demuestra que las transformaciones conformes asociadas con las proyecciones de Lagrange son

$$f(z) = \frac{1}{Ae^{icz} + B} + D\,, \tag{1.2}$$

donde $A, B \in \mathbb{C} \setminus \{0\}$, $D \in \mathbb{C}$ y $c > 0$ (*exponente de la proyección*). En otras palabras, *las proyecciones de Lagrange se obtienen mediante transformaciones racionales lineales de la función exponencial* e^{icz}. También son de Lagrange, en un sentido general, las transformaciones $f(iz)$ donde f está dada por (1.2). En esta nota no consideramos estas transformaciones por las mismas razones expuestas por Lagrange [6, §21]:

> En primer lugar observo que las fórmulas halladas en la primera Memoria (12 y 16) contienen dos soluciones diferentes, ya que se ha visto (11) que está permitido cambiar t por u y u por t; pero observo al mismo tiempo que la solución que resultaría de esta permutación sería más curiosa que cómoda en la práctica, ya que contendría exponenciales del ángulo t y senos y cosenos de la cantidad u que, hemos demostrado, es una cantidad logarítmica.[1] [Aquí la variable t es la longitud λ y u es igual a $-q$.]

Una forma equivalente de escribir la función (1.2) es

$$\begin{aligned} f(z) &= k\frac{1 - e^{ic(z-z_0)}}{1 + e^{ic(z-z_0)}} + w_0 \\ &= -ik \operatorname{tg}\left[\frac{c}{2}(z - z_0)\right] + w_0\,, \end{aligned} \tag{1.3}$$

donde $z_0 \in (-\pi c^{-1}\,,\, \pi c^{-1}) \times \mathbb{R}$ es la preimagen de B/A con respecto a la función e^{icz}, $w_0 = f(z_0) = (2B)^{-1} + D$ y $k = (2B)^{-1}$. Con la introducción de z_0 se consigue que $f(z)$ sea finita en la banda $(-\pi c^{-1}\,,\, \pi c^{-1}) \times \mathbb{R}$. Observamos que $f''(z_0) = 0$, entonces, según el Teorema 2.3, el meridiano y el paralelo de z_0 se transforman en rectas, es decir, z_0 es el centro de la proyección. La última expresión de (1.3), para el caso $\operatorname{Arg} k = 0$, $z_0 = ib$ y $w_0 = 0$, la obtuvo O. Bonnet [2] en su *Thèse d'Astronomie* titulada *Sur la Théorie mathématique des Cartes géographiques*.

Dados $z_0 \in B$ y $c > 0$, definimos entonces el conjunto de las transformaciones bipolares de Lagrange de exponente c y centro z_0 como

$$L(c, z_0) = \left\{ a \operatorname{tg}\left[\frac{c}{2}(z - z_0)\right] + b :\ a, b \in \mathbb{C} \right\}\,, \tag{1.4}$$

y elegimos como representantes canónicos de $L(c, z_0)$ a las transformaciones

$$w = k \operatorname{tg}\left[\frac{c}{2}(z - z_0)\right]$$

[1] Traducción del original de D. Jorge L. Andrés.

con $k > 0$. El propósito de esta nota es obtener una propiedad característica de los representantes canónicos, que expresa la dependencia lineal entre la función $f(z)$ y su derivada pre-schwarziana (véase, por ejemplo, la Introducción en [5]) o derivada logarítmica de $f'(z)$:

$$P_f(z) = \frac{d}{dz} \log f'(z) = \frac{f''(z)}{f'(z)} \,. \tag{1.5}$$

Terminamos esta Introducción con una breve descripción geométrica de las proyecciones de Lagrange. Para simplificar tomamos el representante canónico de $L(c, 0)$ con $k = 1$. En primer lugar, los polos geográficos se transforman en dos puntos finitos y distintos, a saber, $w_n = +i$ y $w_s = -i$. En estos puntos, el ángulo de intersección del meridiano proyectado con el eje de la proyección (segmento $w_s w_n$) es igual a $c\lambda$. En otras palabras, la imagen del meridiano de longitud λ es el arco capaz del segmento $w_s w_n$, de ángulo $c \mid \lambda - \lambda_0 \mid$, Figura 3.

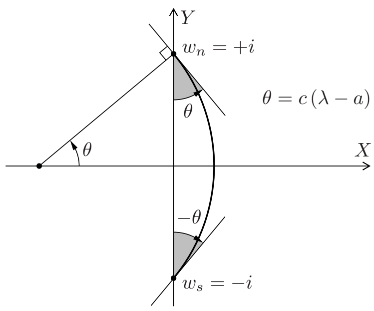


Figura 3: Imagen del meridiano de longitud λ en las proyecciones de Lagrange de exponente c.

Las proyecciones de los paralelos son circunferencias de Apolonio con respecto a w_n y w_s, es decir a lo largo de cualquier paralelo el cociente entre las distancias a los polos se mantiene constante. Esto se deduce de la relación inversa

$$e^{icz} = \frac{i - w}{i + w},$$

que implica $\exp(-cq) = \mid i - w \mid / \mid i + w \mid$.

2. Función característica y derivada pre-schwarziana

El resultado fundamental para obtener las transformaciones conformes de Lagrange es el siguiente [6, § 7]

Teorema 2.1 (Lagrange) *Sea $f(z)$ una función analítica y localmente univalente. Las curvaturas de las imágenes de las rectas verticales y horizontales de la red cartesiana, κ_1 y κ_2 respectivamente, vienen dadas por*

$$\kappa_1 = -\frac{\partial m}{\partial u}\,, \quad \kappa_2 = \frac{\partial m}{\partial v}\,, \tag{2.1}$$

donde $m = |f'(z)|^{-1}$.

Demostración. Sea z un punto del dominio de definición de f, y sean c_1, c_2 las curvas imágenes de las rectas $u = \mathrm{Re(z)}$, $v = \mathrm{Im(z)}$, respectivamente. Escribiendo la primera derivada en forma polar, $f'(z) = |f'(z)| \exp(i\gamma)$, entonces

$$\log f'(z) = -\log m + i\gamma\,. \tag{2.2}$$

Esta función es también analítica y por consiguiente las funciones $-\log m$ y γ satisfacen las condiciones de Cauchy-Riemann:

$$-\frac{\partial \log m}{\partial u} = \frac{\partial \gamma}{\partial v}\,, \quad \frac{\partial \log m}{\partial v} = \frac{\partial \gamma}{\partial u}\,.$$

El ángulo γ, argumento de $f'(z)$, coincide en todo punto, con el ángulo que la tangente a c_2 en $f(z)$ forma con el eje u. Esto se deduce de la identidad $f(\alpha(t))' = f'(\alpha(t))\alpha'(t)$, válida para cualquier curva parametrica diferenciable $\alpha(t)$, pues en el caso de c_2, $\alpha'(t) = 1$ (si la curva es c_1 entonces $\alpha'(t) = i$). Por consiguiente, la derivada de γ con respecto de la longitud de arco de c_2 es la curvatura de c_2, y observando que $ds_2 = |f'(z)|du = m^{-1}du$, tenemos

$$\kappa_2 = \frac{d\gamma}{ds_2} = m\frac{\partial \gamma}{\partial u} = m\frac{\partial \log m}{\partial v}\,.$$

Con esto concluye la demostración para las imágenes de las rectas horizontales. La demostración es análoga para las rectas verticales. □

Lagrange califica a las fórmulas (2.1) como "sencillísimas ecuaciones". En el caso de proyecciones cartográficas conformes, este resultado permite calcular las curvaturas de las imágenes de los meridianos y paralelos a partir de la función $m(z)$ y por esta razón la llamamos *función característica* de la proyección. Si $f(z)$ es una transformación de Lagrange entonces la curvatura de los meridianos solo depende de u, $\kappa_1 = \kappa_1(u)$, y la de los paralelos solo depende de v, $\kappa_2 = \kappa_2(v)$. Tenemos por tanto el siguiente teorema de caracterización de las transformaciones conformes de Lagrange.

Teorema 2.2 (Lagrange) *Una transformación conforme $f(z)$ es de Lagrange si y solo si*

$$\frac{\partial^2 m}{\partial u\,\partial v} = 0\,. \tag{2.3}$$

Un resultado de naturaleza similar lo obtiene Bell [1] para las transformaciones racionales lineales.

También la derivada pre-schwarziana de $f(z)$, definida en la Sección 1, se obtiene a partir de la función característica.

Proposición 2.1 *Sea $f(z)$ una función analítica y localmente univalente, y sea $m(z)$ su función característica. Entonces*

a.
$$P_f(z) = -\frac{\partial \log m}{\partial u} + i\frac{\partial \log m}{\partial v}\,. \tag{2.4}$$

b.
$$\frac{\partial^2 m}{\partial u\,\partial v} = m\,\mathrm{Im}(P_f' - P_f^2/2)\,. \tag{2.5}$$

Demostración. *a.* Partimos de la identidad (2.2),

$$\log f'(z) = -\log m + i\gamma\,,$$

donde γ es el argumento de $f'(z)$. Derivando esta expresión obtenemos

$$\begin{aligned} P_f(z) &= -\frac{\partial \log m}{\partial u} + i\frac{\partial \gamma}{\partial u} \\ &= -\frac{\partial \log m}{\partial u} + i\frac{\partial \log m}{\partial v}\,, \end{aligned}$$

donde hemos aplicado la segunda condición de Cauchy-Riemann al par de funciones $\{-\log m, \gamma\}$.

b. Para simplificar notación escribimos P en lugar de P_f. Derivando (2.4) obtenemos

$$P' = -(\log m)_{uu} + i(\log m)_{uv}\,,$$

y por consiguiente $m^2\,\mathrm{Im}\,P' = m\,m_{uv} - m_u m_v = m\,m_{uv} + m^2\,\mathrm{Im}(P^2/2)$, es decir,

$$m_{uv} = m\,\mathrm{Im}\left(P' - P^2/2\right).$$

□

El apartado (b) de este teorema permite caracterizar a las proyecciones de Lagrange a través de la derivada pre-schwarziana con la condición (véase el Teorema 2.2)

$$P_f' - P_f^2/2 = k \in \mathbb{R}\,, \tag{2.6}$$

donde el primer miembro se conoce como *derivada de Schwarz* de $f(z)$ [4].

Según esta Proposición y el Teorema 2.1 podemos concluir la siguiente identidad que relaciona la derivada pre-schwarziana con las curvaturas fundamentales.

Teorema 2.3 *Sea $f(z)$ una función analítica y localmente univalente, entonces*

$$mP_f(z) = \kappa_1 + i\kappa_2. \quad \blacksquare \tag{2.7}$$

Ejemplo. Sea $f(z) = \operatorname{tg}(cz/2)$ un representante canónico de $L(c, 0)$. Su función característica es

$$m(z) = c^{-1}\left[\cos(cu) + \cosh(cv)\right].$$

Entonces las curvaturas fundamentales son:

$$\kappa_1 = \operatorname{sen}(cu), \quad \kappa_2 = \operatorname{senh}(cv),$$

y la derivada pre-schwarziana es

$$\begin{aligned} P_f(z) &= c \operatorname{tg}(cz/2) \\ &= c\left[\frac{\operatorname{sen}(cu)}{\cos(cu) + \cosh(cv)} + i\frac{\operatorname{senh}(cv)}{\cos(cu) + \cosh(cv)}\right]. \end{aligned}$$

3. Representantes canónicos de las proyecciones de Lagrange

Al comienzo de su segunda memoria, Lagrange [6, § 21] recapitula y presenta de manera simplificada las ecuaciones de sus proyecciones. Dichas ecuaciones se corresponden con lo que aquí hemos llamado representantes canónicos de las proyecciones de Lagrange. De estas ecuaciones se deduce que

$$c\,f(z) = m^{-1}(\kappa_1 + i\kappa_2). \tag{3.1}$$

Según el Teorema 2.3 observamos que se cumple la identidad $c\,f(z) = P_f(z)$. Por tanto esta importante propiedad está implícita en la segunda memoria de Lagrange. El recíproco también es cierto, y podemos formular el siguiente teorema de caracterización de los representantes canónicos de las proyecciones de Lagrange. Utilizamos la siguiente notación: llamamos $T(B, z_0)$ al conjunto de transformaciones conformes de B tales que en el punto $z_0 \in B$ satisfacen las condiciones $f(z_0) = 0$, $\operatorname{Arg} f'(z_0) = 0$ y $f''(z_0) = 0$.

Teorema 3.1 *Sea $z_0 \in B$ y $f(z)$ una función de $T(B, z_0)$.*

a. Si $f(z)$ es una transformación bipolar de Lagrange, entonces P_f y f son linealmente dependientes, es decir, existe una constante $K > 0$, tal que

$$P_f(z) = Kf(z), \tag{3.2}$$

para todo $z \in B$.

b. Si P_f y f son linealmente dependientes, entonces $f(z)$ es un representante canónico de las proyecciones de Lagrange con centro en z_0.

Demostración. *a.* Para simplificar notación escribimos P en lugar de P_f. Según la ecuación (2.6), la derivada pre-schwarziana de $f(z)$ es solución de la ecuación diferencial

$$P' - P^2/2 = k,$$

sujeta a la condición inicial $P(z_0) = 0$, donde k es real. Suponiendo $k \neq 0$, la solución de este primer problema es

$$P = \sqrt{2k}\,\mathrm{tg}\left(\sqrt{k/2}\,z'\right), \tag{3.3}$$

donde $z' = z - z_0$. Conocida la pre-schwarziana, para determinar $f'(z)$ hay que integrar la ecuación diferencial

$$\frac{d}{dz}\log f'(z) = P,$$

con la condición $\operatorname{Arg} f'(z_0) = 0$. Se obtiene

$$f'(z) = \frac{k'}{\cos^2\left(\sqrt{k/2}\,z'\right)},$$

donde $k' > 0$. Finalmente tenemos

$$f(z) = k'\sqrt{2/k}\,\mathrm{tg}\left(\sqrt{k/2}\,z'\right), \tag{3.4}$$

una vez aplicada la última condición $f(z_0) = 0$. Si comparamos (3.3) y (3.4) concluimos que $P = (k/k')f$. La demostración concluye suponiendo $k > 0$, por ejemplo, $k = c^2/2$ con $c > 0$, pues en este caso la función $f(z)$ se escribe

$$f(z) = (2k'/c)\,\mathrm{tg}\left(cz'/2\right),$$

representante canónico de $L(c, z_0)$.

El caso $k < 0$ equivale a $f(iz)$; el caso $k = 0$ equivale a la ecuación $2P' - P^2 = 0$, que conduce a las transformaciones racionales lineales, también excluidas en nuestro estudio.

b. La condición (3.2) es equivalente a la ecuación diferencial

$$f' - (K/2)f^2 = A,$$

donde A es una constante compleja. Necesariamente $A > 0$, porque $f(z_0) = 0$ y $\operatorname{Arg} f'(z_0) = 0$. La solución de esta ecuación de Riccati con la condición $f(z_0) = 0$ es

$$f(z) = (c/k)\,i\,\frac{1 - e^{icz'}}{1 + e^{icz'}} = (c/k)\,\mathrm{tg}\left(cz'/2\right),$$

donde $c = \sqrt{2KA}$. Queda así terminada la demostración del teorema. □

Referencias

[1] P.O. Bell: *A characterization of the group of homographic transformations.* Bull. Amer. Math. Soc. **47**(6), (1941), 488–493.

[2] O. Bonnet: *Thèse d'Astronomie. Sur la Théorie mathématique des Cartes géographiques.* J. Math Pures Appl **17**(1), (1852), 301–340.

[3] P.L. Chebyshev (1856): *Sur la construction des cartes géographiques.* In: *Ouvres I,* (1962), Chelsea, NewYork, 233–236 y 239–247.

[4] P.L. Duren: *Univalent functions.* Springer-Verlag, New York, 1983.

[5] R. Hernández, M.J. Martín: *Pre-Schwarzian and Schwarzian Derivatives of Harmonic Mappings.* J. Geom. Anal. **25**(1) (2015) 64–91.

[6] J.L. Lagrange: *Sur la construction des cartes géographiques.* Nouveaux Mémoires de L'académie Royale Des Sciences Et Belles-Lettres, année 1779, Berlin, Premier Mémoire: 161–185, Second Mémoire: 186–210. También en: *Œuvres de Lagrange*, tomos 1 a 14, Gauthier-Villars, Paris, 1867–1892. **Tomo 4**, (1869), 637–692.

[7] J.P. Snyder, P.M. Voxland: *An album of map projections.* U.S. Geological Survey professional paper 1453, Washington D. C., 1989.

[8] N.A. Tissot: *Mémoire sur la représentation des surfaces et les projections des cartes géographiques.* Gauthier-Villars, Paris, 1881.

The dodecahedron: from intersections of quadrics to Borromean rings

Enrique Artal Bartolo[*], **Santiago López de Medrano**[†] **and María Teresa Lozano**[‡]

IUMA, Universidad de Zaragoza, 50009 Zaragoza, Spain, artal@unizar.es

Instituto de Matemáticas, Universidad Nacional Autónoma de México, 04510 México, D.F., México, santiago@matem.unam.mx

IUMA, Universidad de Zaragoza, 50009 Zaragoza, Spain, tlozano@unizar.es

Dedicado a nuestro amigo José María Montesinos Amilibia con gratitud y afecto.

ABSTRACT

The goal of this paper is to study from different points of view a manifold Z_D which is associated to a regular dodecahedron. It is a real moment-angle manifold constructed from the dodecahedron; we provide algebraic equations of this manifold using only polynomials of degree 2. This manifold has a hyperbolic manifold structure tesselated by right-angled hyperbolic dodecahedra. We study its relationship with others hyperbolic orbifolds related to the dodecahedron, and we obtain, among others results, the orbifold covering from Z_D to the hyperbolic orbifold structure on the 3-sphere with right-angled singularities at the Borromean rings.

2010 Mathematics Subject Classification: 57M10, 57M50
Key words: Orbifold, dodecahedron, orbifold covering, quadrics.

[*]Author partially supported by project MTM2013-45710-C2-1-P and DGA Grupo Consolidado E15.

[†]Author partially supported by project IN111415-PAPIIT and CEI Campus Iberus

[‡]Author partially supported by project MTM2013-45710-C2-1-P and DGA Grupo Consolidado E15.

Introduction

In the first term of the academic year 2014/15, the second named author of this paper gave a course on *Topology of intersection of quadrics* in the `Seminario de Geometrıa y Topología` organized by the IUMA at the Universidad de Zaragoza.

Among the many interesting and fruitful discussions around this course, one was related with the topological and geometric structure of the low-dimensional manifolds (or pseudo-manifolds) appearing in this context. The surfaces that appear as such intersections are well-known and classified, see Example 1.1. So, our interest turned out immediately to the 3-dimensional case.

Intersection of quadrics appear also as real moment-angle spaces associated to polytopes. For the 2-dimensional case the n-polygon is studied, while for the 3-dimensional case the starting object are polyhedra. These polyhedra have to be simple to yield smooth manifolds but the singular case is also interesting. If we restrict our attention to regular polyhedra, only the tetrahedron, the cube and the dodecahedron yield manifolds. The two first cases are easily studied, see Example 1.2, and our attention was directed to the dodecahedron D.

The manifold Z_D associated to a regular dodecahedron can be described in several ways. It is defined as an intersection of 9 diagonal quadrics in $\mathbb{R}^{12}$, see (1.3); the quotient of $Z_{\mathbf{D}}$ by the reflections along all the coordinate hyperplanes is isomorphic to $Z_D \cap \mathbb{R}^{12}_{+}$ which can be linearized to obtain the dodecahedron D. From this dodecahedron D, we can recover the manifold by reflecting it along all the coordinate hyperplanes. In this way the dodecahedron acquires an orbifold structure $\mathbf{D}$ with mirror faces, which is actually hyperbolic with dihedral angles equal to $\frac{\pi}{2}$ and we will denote the hyperbolic manifold as $Z_{\mathbf{D}}$. The manifold $Z_{\mathbf{D}}$ is tesselated by 2^{12} dodecahedra. This tesselation has nice symmetry properties (it is super-regular).

There is another important hyperbolic orbifold related to the dodecahedron, the universal orbifold $B_{4,4,4}$ whose underlying topological space is the 3-sphere and the $\frac{\pi}{2}$-singularities are located at the Borromean rings and whose fundamental group U has a universal property [HLMW87]. We relate $Z_{\mathbf{D}}$ and $B_{4,4,4}$ through orbifold coverings, see (3.9). As a consequence, $\pi_1(Z_{\mathbf{D}})$ is an index-2^{12} torsion-free subgroup of the universal group U.

The paper is organized as follows. In §1 we introduce the relationship between polyhedra and intersection of quadrics and as a consequence we give algebraic equations for the manifold $Z_{\mathbf{D}}$, where all of them are of degree 2. In §2 we analyze the hyperbolic structure of the orbifold $\mathbf{D}$ and the manifold $Z_{\mathbf{D}}$. This study allows to prove that $Z_{\mathbf{D}}$ is orientable, admits a super-regular tesselation and covers the well-known Löbel manifold (which itself covers $\mathbf{D}$). Finally, in §3 we relate $Z_{\mathbf{D}}$ to the universal orbifold $B_{4,4,4}$, and we relate the universal abelian covers of the above orbifolds.

We will use the following notations:

- For $n \in \mathbb{N}$, C_n is the cyclic group of order n.

- For $n \in \mathbb{N}$, Σ_n is the symmetric group of n elements.
- For $n \in \mathbb{N}$, $[n]$ will denote the set $\{1, 2, \ldots, n\}$
- $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$.
- The golden ratio $\frac{1+\sqrt{5}}{2}$ will be denoted by ϕ.

1. Intersections of quadrics and polytopes

Intersection of quadrics are related to polytopes. Let A be a $k \times n$ matrix of rank k with entries in $\mathbb{R}$ and let $A_i \in \mathbb{R}^k$ be its columns. We denote by $V = V(A)$ the intersection of the quadrics in $\mathbb{R}^n$ given by the equations

$$\sum_{i=1}^{n} A_i X_i^2 = 0 \tag{1.1}$$

and by $Z = Z(A)$ the intersection of V with the unit sphere $\sum_{i=1}^{n} X_i^2 = 1$. Let also $\Pi = \Pi(A)$ be the affine subspace of $\mathbb{R}^n$ given by

$$\sum_{i=1}^{n} A_i X_i = 0, \quad \sum_{i=1}^{n} X_i = 1,$$

and $P = P(A)$ be the convex polytope $\Pi(A) \cap (\mathbb{R}_+)^n$. Both Z and P have dimension $d = n - k - 1$. The polytope P is homeomorphic to $Z \cap (\mathbb{R}_+)^n$ and, topologically, Z can be recovered from P by reflecting it in all coordinate hyperplanes.

Generically, the polytope P is transversal to all faces of $\mathbb{R}^n_+$. In that case, it is easy to see that Z is a smooth variety and P is a simple polytope[1]. In the transverse case, Z is completely determined by the combinatorics of P as a quotient of $P \times C_2^n$, while in general this may depend on the way P is embedded in $\mathbb{R}^n_+$.

It is well-known that any convex polytope of dimension d with m facets can be realized as $P(A) \subset R^n$ for any $n \geq m$ and some $k \times n$ matrix A of rank $k = n - d - 1$. Further, if the polytope is simple it can be realized transversely.

Example 1.1 Let us consider an n-polygon P_n in $\mathbb{R}^n_+$ such that each edge is the intersection with one hyperplane coordinate. The manifold Z associated to P_n is an orientable Riemann surface of genus $g_n = 2^{n-3}(n-4) + 1$, see [LdM14].

Example 1.2 There are three simple regular polytopes in dimension 3. If P is a tetrahedron it is easy to prove that the corresponding manifold Z is actually $\mathbb{S}^3 \subset \mathbb{R}^4$. Let us embed the cube Q in $\mathbb{R}^6$. We start with the symmetric cube in $\mathbb{R}^3$, i.e. the convex hull of its vertices $(\pm 1, \pm 1, \pm 1)$. The normal vectors N_i of its faces are

[1] A convex polytope of dimension d is *simple* if every one of its vertices lies in exactly d facets.

$\pm\mathbf{e}_i$ ($\mathbf{e}_i$ is the i^{th} vector of the canonical basis. Each face is in the affine plane $\{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{x} \cdot N_i = 1\}$. The affine map $\mathbb{R}^3 \to \mathbb{R}^6$ given by

$$\mathbf{x} \mapsto (X_1, \ldots, X_6) = (1 - \mathbf{x} \cdot \mathbf{e}_1, 1 - \mathbf{x} \cdot \mathbf{e}_2, 1 - \mathbf{x} \cdot \mathbf{e}_3, 1 + \mathbf{x} \cdot \mathbf{e}_1, 1 + \mathbf{x} \cdot \mathbf{e}_2, 1 + \mathbf{x} \cdot \mathbf{e}_3)$$

embeds the cube Q as expected. We can check easily that Q is defined by:

$$X_i \geq 0, \quad X_1 + X_4 = 2, \quad X_2 + X_5 = 2, \quad X_3 + X_6 = 2.$$

The equations of the quadrics whose intersection is Z are:

$$X_1^2 + X_4^2 = 2, \quad X_2^2 + X_5^2 = 2, \quad X_3^2 + X_6^2 = 2.$$

From this equations is clear that Z is homeomorphic to $(\mathbb{S}^1)^3$; in fact we obtain the Clifford torus, where the induced metric structure is euclidean. We can rewrite the equations as in (1.1):

$$\sum_{i=1}^{6} X_i^2 = 6, \quad X_1^2 - X_2^2 + X_4^2 - X_5^2 = 0, \quad X_2^2 - X_3^2 + X_5^2 - X_6^2 = 0.$$

We use the sphere of radius $\sqrt{6}$ to obtain the simplest coefficients.

In the smooth case the topology of Z (and other related spaces) has been studied for sometime now ([Wal80, LdM89, LdM14, GLdM13]). Independently, in [DJ91, section 4.1], the construction of Z is given abstractly and identified as the *universal abelian cover* of P viewed as an orbifold. It is sometimes called a *real moment-angle manifold* (see footnote 2 below) and is part of a very general and abstract construction called the polyhedral product functor, see [BBCG10].

The main interest in [DJ91] is the study of other smooth covers of P of order 2^d called *small covers* which are the real topological analogs of the complex projective toric varieties[2]. It is shown there that they do not exist for all polytopes, but that they do exist for 3-dimensional simple ones[3].

In Example 1.2 we skipped the case of the dodecahedron D and its corresponding real moment-angle manifold. This manifold will be the core of this paper. The goal of the rest of this section is to give explicit and symmetric equations of this manifold as intersection of quadrics (1.1); together with the equation of a sphere centered at the origin. As we did in Example 1.2, we will take the sphere of radius $\sqrt{12}$ to obtain the simplest coefficients.

The first step is to embed the regular dodecahedron in $\mathbb{R}^{12}_+$ such that each face F_i is the intersection with the coordinate hyperplane $\{X_i = 0\}$. We number the faces F_i

[2]There are other related manifolds with torus actions called *moment-angle* manifolds and *(quasi)-toric* manifolds

[3]The proof of this fact uses the Four Color Theorem!

of the dodecahedron as in Figure 1. We start with the regular dodecahedron in $\mathbb{R}^3$, whose vertices v_k are

$$\{(\pm 1, \pm 1, \pm 1)\} \cup \bigcup_{j=0}^{2} \{(0, \pm\phi, \pm\phi^{-1})\}^{\rho^j}$$

where ρ stands for the right action $(x, y, z)^\rho := (z, x, y)$. This is the dodecahedron inscribed in the sphere of radius $\sqrt{3}$. The normal vectors N_j of the faces are all the cyclic permutations of $(\pm(2-\phi), \pm(\phi-1), 0)$ ordered as in Figure 1 (we leave to the reader to give an explicit ordering). If a vertex is in a face, it satisfies $N_j \cdot v_k = 1$. The embedding of D in $\mathbb{R}^{12}$ is given by the affine map $\mathbb{R}^3 \to \mathbb{R}^{12}$:

$$(x, y, z) \mapsto (1 - (x, y, z) \cdot N_1, \ldots, 1 - (x, y, z) \cdot N_{12}).$$

The group of isometries of the dodecahedron (as permutation of the faces) is generated by the following permutations of the coordinates in $\mathbb{R}^{12}$:

1. Central symmetry: $(1\,12)(2\,11)(3\,10)(4\,9)(5\,8)(6\,7)$.

2. Rotation of angle $\frac{2\pi}{5}$ around the axis joining the centers of faces 1 and 12: $(2\,3\,4\,5\,6)(7\,11\,10\,9\,8)$

3. Rotation of angle $\frac{2\pi}{3}$ around the axis joining the vertices $(1, 2, 6)$ and $(7, 11, 12)$: $(1\,2\,6)(3\,9\,5)(4\,8\,10)(7\,12\,11)$.

4. Rotation of angle π around the axis joining the centers of the edges $(1, 2)$ and $(11, 12)$: $(1\,2)(3\,6)(4\,9)(5\,8)(7\,10)(11\,12)$.

One of the vertices is:

$$\frac{1}{6}\,(0,\,0,\,0,\,2-\phi,\,\phi-1,\,2-\phi,\,\phi-1,\,2-\phi,\,\phi-1,\,1,\,1,\,1)\,, \tag{1.2}$$

and the other ones are obtained using the symmetry group.

Equalities involving normal vectors N_i imply equations satisfied by the points in the dodecahedron in $\mathbb{R}^{12}$. They are:

1. $N_i + N_{13-i} = 0$, i.e. $X_i + X_{13-i} = 2$. Adding up, we obtain $\sum_{i=1}^{12} X_i = 12$ and the following system of 15 homogeneous $\{X_i + X_{13-i} - X_j - X_{13-j} = 0 \mid 1 \leq i < j \leq 6\}$ which are invariant by the above group of permutations (eventually a permutation can change the sign of the equation).

   Note that only 5 of these equations are linearly independent (say $(i, j) = (1, 2), (2, 3), (3, 4), (4, 5), (5, 6)$).

2. Consider a face, say the first one, and the five neighboring faces. We obtaing the equality $N_2 + \cdots + N_6 - \sqrt{5}N_1 = 0$. Hence, we have:

$$X_2 + \cdots + X_6 - \sqrt{5}X_1 = 5 - \sqrt{5}.$$

As before, we can consider the homogeneous equations associated to a pair of faces. In fact, we can restrict our attention to the equations (up to sign) associated to a pair of adjacent faces, i.e., to edges, denoted as $i \asymp j$. If we denote an edge by its neighboring faces $\{i, j\}$, we have 30 equations:

$$\left\{ \sum_{k \asymp i} X_k - \sqrt{5}X_i = \sum_{\ell \asymp j} X_\ell - \sqrt{5}X_j \mid \{i, j\} \text{ edge} \right\}.$$

These 30 equations contain 8 independent equations, e.g., consider the equations associated to a set E formed by four edges of F_1 and four edges of F_2 such that the forgotten edges are not opposite. Finally, we can give a family of 9 equations of the real moment-angle manifold associated to the dodecahedron:

$$\sum_{i=1}^{12} X_i^2 = 12, \quad \left\{ \sum_{k \asymp i} X_k^2 - \sqrt{5}X_i^2 = \sum_{\ell \asymp j} X_\ell^2 - \sqrt{5}X_j^2 \mid \{i, j\} \in E \right\}. \tag{1.3}$$

2. Orbifold covering spaces of the right angle dodecahedron

In the previous section we studied some properties of the regular dodecahedron $D \subset \mathbb{R}^3$ and of its embedding in $\mathbb{R}^{12}$. In this section we study the hyperbolic regular dodecahedron $\mathbf{D}$ such that every dihedral angle defined by two adjacent faces is $\frac{\pi}{2}$. The hyperbolic space H^3 has a super regular tessellation T by copies of $\mathbf{D}$ [Thu80]. The group $G_{\mathbf{D}}$ generated by reflection on the twelve planes containing the faces of any dodecahedron acts on H^3 preserving the tessellation and has the dodecahedron as fundamental domain. The quotient $H^3/G_{\mathbf{D}}$ defines a hyperbolic orbifold structure in $\mathbf{D}$ with mirror singularity on every face. Then the reflection group $G_{\mathbf{D}}$ is the orbifold fundamental group of $\mathbf{D}$. In order to understand the group $G_{\mathbf{D}}$ we number the faces of $\mathbf{D}$ as in Figure 1. Let x_i be the reflection on the face $\mathbf{i}$. If $\mathbf{i}$ and $\mathbf{j}$ are two adjacent faces, denote by l_{ij} the common edge. The vertex v_{ijk} is the common vertex of $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ when they intersect. The group $G_{\mathbf{D}}$ is generated by $x_1, ..., x_{12}$. The isotropy subgroups $H(\bullet)$ of the singular elements are given by the following generators and relations:

$$\begin{array}{lcl} H(\mathbf{i}) & = & |x_i;\, x_i^2| = C_2 \\ H(l_{ij}) & = & |x_i, x_j;\, x_i^2, x_j^2, (x_i x_j)^2| = C_2 \times C_2 \\ H(v_{ijk}) & = & |x_i, x_j, x_k;\, x_i^2, x_j^2, x_k^2, (x_i x_j)^2, (x_j x_k)^2, (x_k x_i)^2| = C_2 \times C_2 \times C_2 \end{array}$$

Hence,

$$G_{\mathbf{D}} = \left| x_1, ..., x_{12} : \underset{1 \leq i \leq 12}{x_i^2}, \quad \underset{l_{ij} \text{ edge}}{(x_i x_j)^2} \right| .$$

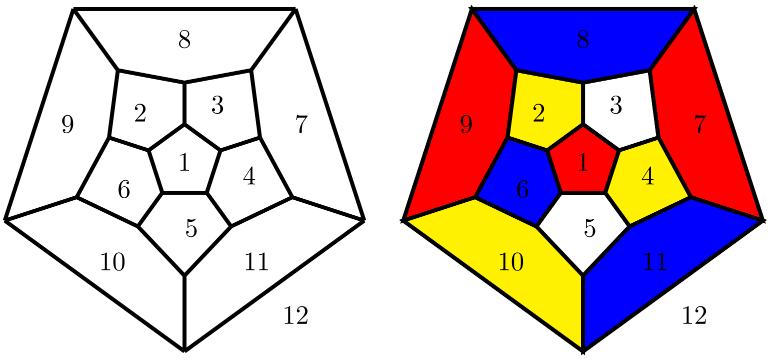


Figure 1: Dodecahedron $\mathbf{D}$ and one coloring on $\mathbf{D}$

The orbifold coverings over $\mathbf{D}$ are classified by the conjugation classes of subgroups of $G_{\mathbf{D}}$. To construct a n-fold covering, we should define a monodromy map on the permutation group of n elements:

$$\omega_n : G_{\mathbf{D}} \longrightarrow \Sigma_n$$

If each permutation $\omega_n(x_i)$, $i = 1, ..., 12$, is the product of $\frac{n}{2}$ different transpositions, then the cover has no mirror singular set. Observe that in this case the number of sheets should be even. The minimal number of sheets to obtain a manifold is at least 8 because the isotropy subgroup of each vertex of $\mathbf{D}$ for the action of $G_{\mathbf{D}}$, is isomorphic to the 8-element abelian group C_2^3. Therefore one needs 8 copies of the fundamental domain around the preimage of each vertex to kill the singularity. For $n = 2, 4, 6$ the cover has singularities. For instance, the case $n = 2$ produces the double of $\mathbf{D}$, that is, an orientable hyperbolic orbifold structure in S^3 with a singular trivalent graph, the 1-skeleton of the dodecahedron, where the angle around the edges is π.

2.1. The Löbell manifold

The Löbell manifold $L(5)$ constructed in [Ves98], see also [Löb31], is an $8:1$ cover of $\mathbf{D}$ defined by the coloring depicted in Figure 1. The procedure to construct the monodromy of the covering is the following: Color the faces with 4 different colors (R, Y, B, W) such that the three faces meting in a vertex have different color. Then define a map sending the generator x_i to an element of C_2^3 according to the assigned color as follows. Let

$$C_2 \times C_2 \times C_2 \equiv (C_2\langle R\rangle \oplus C_2\langle Y\rangle \oplus C_2\langle B\rangle \oplus C_2\langle W\rangle) \,/\, (C_2\langle R+Y+B+W\rangle)\,.$$

The map ρ_L

$$\begin{aligned}
\rho_L : G_{\mathbf{D}} &\longrightarrow C_2 \times C_2 \times C_2\\
x_1, x_7, x_9 &\longmapsto R \equiv (1,0,0)\\
x_2, x_4, x_{10} &\longmapsto Y \equiv (0,1,0)\\
x_6, x_8, x_{11} &\longmapsto B \equiv (0,0,1)\\
x_3, x_5, x_{12} &\longmapsto B \equiv (1,1,1)
\end{aligned}$$

is a homomorphism because each generator goes to an order two element and the group C_2^3 is abelian. On the other hand, the three different colors at each vertex is a necessary and sufficient condition in order to have a manifold.

Now number the eight elements of $C_2 \times C_2 \times C_2$, for instance:

$$\begin{array}{llll}
(1,0,0)=1 & (0,0,0)=2 & (0,1,0)=3 & (1,1,0)=4\\
(0,0,1)=5 & (1,0,1)=6 & (0,1,1)=7 & (1,1,1)=8
\end{array}$$

and define the monodromy as the permutation associated to the left action of $\rho(x_i)$ onto C_2^3:

$$\begin{aligned}
\omega_L : G_{\mathbf{D}} &\longrightarrow \Sigma_8\\
x_1, x_7, x_9 &\longmapsto (12)(34)(56)(78)\\
x_2, x_4, x_{10} &\longmapsto (14)(23)(57)(68)\\
x_6, x_8, x_{11} &\longmapsto (16)(25)(37)(48)\\
x_3, x_5, x_{12} &\longmapsto (17)(28)(36)(45)
\end{aligned}$$

This monodromy ω_L define the 8-fold orbifold covering

$$p_{L(5)} : L(5) \xrightarrow{8:1} \mathbf{D}$$

The fundamental group of the manidold $L(5)$ is the preimage by ω_L of the stabilizer in Σ_8 of one element, for instance 1. By the above construction of ω_L it is easy to see that this subgroup is the kernel of the homomorphism ρ_L. The kernel is a normal subgroup, then the orbidold covering is regular and the group of deck transformations is C_2^3.

This procedure can be generalized to construct the monodromy of some 2^s-fold coverings ($4 \le s \le 12$) as follows: The number n of different colors on the faces should be 4, 6 or 12 colors such that the three faces meting in a vertex have different color and the number of faces of the same color coincides (3, 2 or 1). The map $\rho : G_{\mathbf{D}} \longrightarrow C_2^s$ should be onto, therefore $n \ge s$.

2.2. The abelian universal covering

We are interested in one of these coverings, the one with more sheets, the abelian universal covering. Here $s = n = 12$. Observe that the homomorphism ρ_u is the abelianization of the group $G_{\mathbf{D}}$.

$$\begin{aligned} \rho_u : G_{\mathbf{D}} &\longrightarrow (C_2)^{12} \\ x_i &\longmapsto (0, \ldots, \underset{(i)}{1}, \ldots, 0) \end{aligned}$$

Then the kernel of ρ_u is the commutator subgroup or derived subgroup $G'_{\mathbf{D}}$, which is the smallest normal subgroup such that the quotient group of the original group $G_{\mathbf{D}}$ by this subgroup is abelian.

As before, the action of $\rho_u(x_i)$ on the 2^{12} elements of $(C_2)^{12}$ defines the monodromy

$$\omega_u : G_{\mathbf{D}} \longrightarrow \Sigma_{2^{12}}$$

of the regular orbifold covering

$$p_u : Z_{\mathbf{D}} \xrightarrow{2^{12}:1} \mathbf{D}$$

The fundamental group of the hyperbolic manifold $Z_{\mathbf{D}}$ is the derived subgroup $G'_{\mathbf{D}}$ and the group of deck transformations is $(C_2)^{12}$.

Theorem 2.1 *The hyperbolic manifold $Z_{\mathbf{D}}$ is a 2^9-fold covering of $L(5)$.*

Proof. The homomorphism ρ_L factors though the homomorphism ρ_u as follows:

$$\begin{tikzcd} G_{\mathbf{D}} \arrow[rr, "\rho_L"] \arrow[dr, "\rho_u"'] & & (C_2)^3 \\ & (C_2)^{12} \arrow[ur, "h"'] & \end{tikzcd}$$

□

2.3. The super regular tessellation of $Z_{\mathbf{D}}$

The hyperbolic manifold $Z_{\mathbf{D}}$ has a super regular tessellation by dodecahedra, because the 2-skeleton and the 1-skeleton also have a regular tessellation by hyperbolic pentagons and geodesics respectively. Next we analyze these submanifolds.

Let l be an edge of the dodecahedron. Then $p_u^{-1}(l)$ is the disjoint union of 2^8 closed geodesics composed by 4 segments of the same length. Therefore the 1-skeleton of $Z_{\mathbf{D}}$ is a 6-valent graph composed by the union of $30 \times 2^8 = 15 \times 2^9$ isometric closed hyperbolic geodesics arranged in 30 families, such that two geodesics in the same family do not intersect, there are exactly 5×2^{11} triple intersection points, and every closed geodesic has 4 of them.

Let P be a face of the dodecahedron. Then $p_u^{-1}(P)$ is the disjoint union of $2^6 = 64$ hyperbolic surfaces of genus 5, F_5. The 2-skeleton of $Z_{\mathbf{D}}$ is the union of $12 \times 2^6 = 3 \times 2^8$ hyperbolic surfaces F_5 arranged in 12 families, such that two surfaces in the same family do not intersect. If two surfaces intersect they do it along a closed geodesic and the surfaces correspond to two adjacent pentagons in the fundamental dodecahedron. Then every surface intersects with surfaces in five other families.

Theorem 2.2 *Every F_5 surface in the preimage by p_u of a face of the dodecahedron is a embedded incompressible surface in $Z_{\mathbf{D}}$.*

Proof. A orientable surface embedded in a manifold is incompressible if the fundamental group of the surface injects in the fundamental group of the manifold. Suppose that F_5 is the surface in the preimage of the face **1** generated by the reflection on the five adjacent pentagons **2**, **3**, **4**, **5** and **6**, see Figure 1. The group $G_{\mathbf{1}}$ of the orbifold structure in **1** is the subgroup of $G_{\mathbf{D}}$ generated by the reflection $(x_2, x_3, x_4, x_5, x_6)$ and F_5 is the abelian universal covering of the orbifold **1**. Then the following diagram is commutative.

$$\begin{array}{ccc} G_{\mathbf{1}} & \xrightarrow{\rho_{u1}} & (C_2)^5 \\ \big\downarrow & & \big\downarrow \\ G_{\mathbf{D}} & \xrightarrow[\rho_u]{} & (C_2)^{12} \end{array}$$

where the vertical maps are inclusions. Therefore the fundamental group of F_5, being the kernel of ρ_{u1}, injects in the kernel of ρ_u, which is the fundamental group of $Z_{\mathbf{D}}$. □

The volume of **D** has been computed in [Ves98] (see also [Ves10, Theorem 3.2]).

Corollary 2.3 *The volume of $Z_{\mathbf{D}}$ is $2^{12}\,\mathrm{Vol}(\mathbf{D}) = 2^9\,\mathrm{Vol}(L(5))$.*

3. The orbifolds D and $B_{4,4,4}$

The hyperbolic orbifold structure $B_{4,4,4}$ in S^3 with singular set the Borromean rings with cyclic isotropy group of order 4 is the quotient of the hyperbolic space H^3 by the universal group U [HLMW87] and has also a regular right angle hyperbolic dodecahedron as fundamental domain ([Thu80]), see Figure 2.

Here is a presentation of U, the fundamental group of the orbifold $B_{4,4,4}$:

$$\begin{aligned} U = \langle a, b, c \mid a\, b\bar{c}\bar{b}c &= b\bar{c}\bar{b}c\, a,\ a^4, \\ b\, c\bar{a}\bar{c}a &= c\bar{a}\bar{c}a\, b,\ b^4, \\ c\, a\bar{b}\bar{a}b &= a\bar{b}\bar{a}b\, c,\ c^4 \,\rangle \end{aligned} \tag{2.1}$$

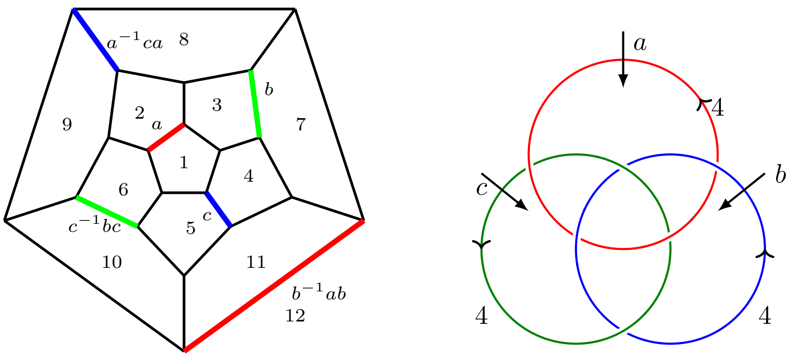


Figure 2: The orbifold $B_{4,4,4}$

The generators a, b, and c arise from the three meridian generators for the three components of the Borromean rings and they are $\frac{\pi}{2}$-rotation around the corresponding edge of the dodecahedron.

The fundamental orbifold group of $B_{4,4,4}$ is U. Therefore both groups, $G_{\mathbf{D}}$ and U act on the hyperbolic space H^3 by isometries fixing the same tessellation T by regular right angle hyperbolic dodecahedra and having the same fundamental domain. Then the two hyperbolic orbifolds $\mathbf{D}$ and $B_{4,4,4}$ have the same volume. They are related as follows.

Depicted in Figure 3 is the intersection of the right angle hyperbolic dodecahedron centered in the origin of coordinates with the positive octant. Let us call $\mathbf{Q}$ the orbifold structure on this hyperbolic polyhedron with mirror faces. The dihedral angle are all right angles, excepted the colored a, b end c which are $\frac{\pi}{4}$. The orbifold fundamental group is generated by the reflection on its six numerated faces. Let us call y_i the reflection on the i face.

Lemma 3.1 *The orbifold* $\mathbf{D}$ *and the orbifold* $B_{4,4,4}$ *are eight-fold regular orbifold covers of the orbifold* $\mathbf{Q}$.

Proof. The monodromy for the covering $q_{\mathbf{D}} : \mathbf{D} \longrightarrow \mathbf{Q}$ is obtained using the homomorphism

$$\begin{aligned} \rho_{DQ} : G_{\mathbf{Q}} &\longrightarrow C_2 \times C_2 \times C_2 \\ y_1 &\longmapsto (1,0,0) \\ y_2 &\longmapsto (0,1,0) \\ y_3 &\longmapsto (0,0,1) \\ y_4, y_5, y_6 &\longmapsto (0,0,0) \end{aligned} \tag{3.1}$$

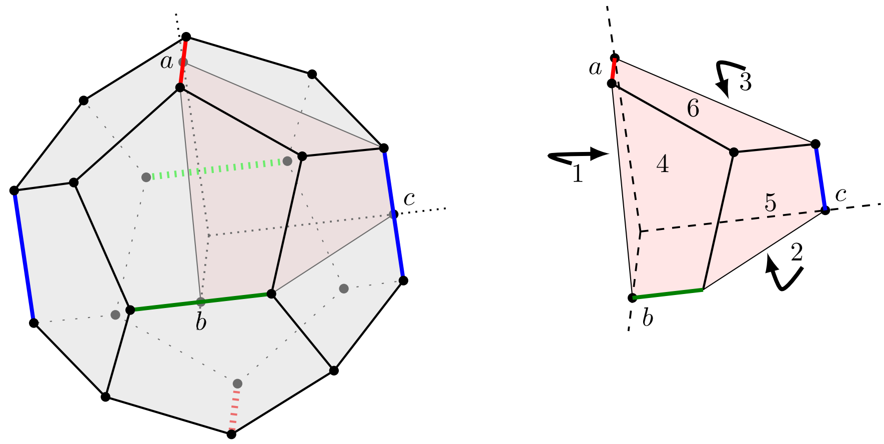


Figure 3: The orbifolds $\mathbb{D}$ and $\mathbf{Q}$.

The monodromy for the covering $q_B : B_{4,4,4} \longrightarrow \mathbf{Q}$ is obtained using the homomorphism

$$\begin{aligned} \rho_{BQ} : G_{\mathbf{Q}} &\longrightarrow C_2 \times C_2 \times C_2 \\ y_1, y_6 &\longmapsto (1,0,0) \\ y_2, y_4 &\longmapsto (0,1,0) \\ y_3, y_5 &\longmapsto (0,0,1) \end{aligned} \tag{3.2}$$

□

Therefore the fundamental groups $G_{\mathbf{D}}$ and U of the orbifolds $\mathbf{D}$ and $B_{4,4,4}$ are index eight subgroups of $G_{\mathbf{Q}}$. Let us recall the concept of commensurable subgroups. Two subgroups A and B of a group are *commensurable* when their intersection has finite index in each of them. This property is an equivalence relation. Therefore, $G_{\mathbf{Q}}$, $G_{\mathbf{D}}$ and U are commensurable subgroups. But we can analyze more commensurable subgroups of $G_{\mathbf{Q}}$.

The universal abelian cover $p_{uQ} : Z_{\mathbf{Q}} \xrightarrow{2^6} \mathbf{Q}$ is associated to the abelianization homomorphism

$$\begin{aligned} \rho_{uQ} : G_{\mathbf{Q}} &\longrightarrow C_2^6 \\ y_i &\longmapsto (0, \ldots, 0, \underset{(i)}{1}, 0, \ldots, 0) \end{aligned} \tag{3.3}$$

Observe that $Z_{\mathbf{Q}}$ is a compact hyperbolic orbifold with singular set composed by closed geodesics of order 2, corresponding to the edges of $\mathbf{Q}$ with angle $\frac{\pi}{4}$.

The universal abelian cover $p_{uB} : Z_B \xrightarrow{2^6} B_{4.4,4}$ is associated to the abelianization homomorphism

$$\begin{aligned} \rho_{uB} : U &\longrightarrow C_4^3 \\ a &\longmapsto (1,0,0) \\ b &\longmapsto (0,1,0) \\ c &\longmapsto (0,0,1) \end{aligned} \tag{3.4}$$

Here Z_B is a compact hyperbolic manifold, all singularities are removed.

The abelian covers are associated to the derived group of the fundamental group of the base orbifold. The commutative diagram (3.5) relates those groups and their derived subgroups. All the maps are inclusions and the label is the index of the subgroup.

$$\begin{tikzcd} G'_{\mathbf{D}} \arrow[rr, "2^{12}"] \arrow[dr, "2^9"] & & G_{\mathbf{D}} \arrow[dr, "2^3"] & \\ & G'_{\mathbf{Q}} \arrow[rr, "2^6"] & & G_{\mathbf{Q}} \\ U' \arrow[ur, "2^3"] \arrow[rr, "4^3"] & & U \arrow[ur, "2^3"] & \end{tikzcd} \tag{3.5}$$

This diagram has a counter-part for covering maps:

$$\begin{tikzcd} Z_{\mathbf{D}} \arrow[rr, "2^{12}"] \arrow[dr, "2^9"] & & \mathbf{D} \arrow[dr, "2^3"] & \\ & Z_{\mathbf{Q}} \arrow[rr, "2^6"] & & \mathbf{Q} \\ Z_B \arrow[ur, "2^3"] \arrow[rr, "4^3"] & & B_{4,4,4} \arrow[ur, "2^3"] & \end{tikzcd}$$

It is natural to look for a common finite cover of the two hyperbolic orbifolds $\mathbf{D}$ and $B_{4,4,4}$. For instance, the smaller one is the covering associated to the intersection of their fundamental groups $H = G_{\mathbf{D}} \cap U$.

Lemma 3.2 *The index of the subgroup H in $G_{\mathbf{D}}$ and U is eight.*

Proof. The subgroup H is the kernel of the surjective homomorphism

$$(\rho_{DQ}, \rho_{BQ}) : G_{\mathbf{Q}} \longrightarrow (C_2)^3 \times (C_2)^3 \tag{3.6}$$

where ρ_{DQ} and ρ_{BQ} are defined in (3.1) and (3.2). Observe that (ρ_{DQ}, ρ_{BQ}) is equal to ρ_{uQ}, see (3.3). The kernel of this homorphism is $G'_{\mathbf{Q}} = H$, has index 2^6 in $G_{\mathbf{Q}}$ and is contained in the index 2^3 subgroups $G_{\mathbf{D}}$ and U which have index 2^3. Therefore the index of H in $G_{\mathbf{D}}$ and U is $2^3 = 8$. $\square$

Theorem 3.3 *The orbifold $Z_{\mathbf{Q}}$ is the minimal common orbifold covering of the hyperbolic orbifolds $\mathbf{D}$ and $B_{4,4,4}$.*

Proof. The homomorphism ρ_{BQ} factors through $\rho_{uQ} = (\rho_{DQ}, \rho_{BQ})$ and so does ρ_{DQ}. Therefore the covering $p_{uQ} : Z_{\mathbf{Q}} \xrightarrow{2^6} \mathbf{Q}$ factors through $q_B : B_{4,4,4} \longrightarrow \mathbf{Q}$ and also through $q_{\mathbf{D}} : \mathbf{D} \longrightarrow \mathbf{Q}$.

$$\begin{tikzcd}
& Z_{\mathbf{Q}} \arrow[dl, "2^3:1"', "p_1"] \arrow[dr, "2^3:1", "p_2"'] \arrow[dd, "\rho_{uQ}"', "2^6:1"] & \\
B_{4,4,4} \arrow[dr, "q_B", "2^3:1"'] & & \mathbf{D} \arrow[dl, "q_{\mathbf{D}}"', "2^3:1"] \\
& \mathbf{Q} &
\end{tikzcd}$$

In fact, we can construct directly the coverings $p_1 : Z_{\mathbf{Q}} \longrightarrow B_{4,4,4}$ and $p_2 : Z_{\mathbf{Q}} \longrightarrow \mathbf{D}$ as follows. The subgroup H_1 generated by the following elements is a subgroup of $H = G_{\mathbf{D}} \cap U$:

$$\begin{array}{lll}
a^2 = x_1x_2, & b^2 = x_3x_9, & c^2 = x_4x_5, \\
b^{-1}a^2b = x_{10}x_{12}, & c^{-1}b^2c = x_6x_{11}, & a^{-1}c^2a = x_7x_7
\end{array}$$

The homomorphim

$$\begin{aligned}
\rho : U &\longrightarrow C_2 \times C_2 \times C_2 \\
a &\longmapsto (1,0,0) \\
b &\longmapsto (0,1,0) \\
c &\longmapsto (0,0,1)
\end{aligned} \tag{3.7}$$

defines the monodromy

$$\begin{aligned}
\omega_1 : U &\longrightarrow \Sigma_8 \\
a &\longmapsto (1\,2)(3\,4)(5\,6)(7\,8) \\
b &\longmapsto (1\,4)(2\,3)(5\,7)(6\,8) \\
c &\longmapsto (1\,6)(2\,5)(3\,6)(4\,8)
\end{aligned}$$

It defines the regular 8-fold orbifold covering of $B_{4,4,4}$

$$p_1 : Z_{\mathbf{Q}} \longrightarrow B_{4,4,4}$$

The map p_1 can be viewed in different ways:

- An 8-fold locally cyclic covering of the sphere S^3 branched over the Borromean rings with branching index 2.

- An 8-fold orbifold covering of $B_{2,2,2}$, where $B_{2,2,2}$ is the Euclidean orbifold structure in S^3 with singular set de Borromean ring with cyclic isotropy group of order 2.

This implies that $Z_{\mathbf{Q}}$ has a Euclidean manifold structure, with no singularity and also a hyperbolic orbifold structure with a singular link. Note that the monodromy in (3.2) factors through the one in (3.4).

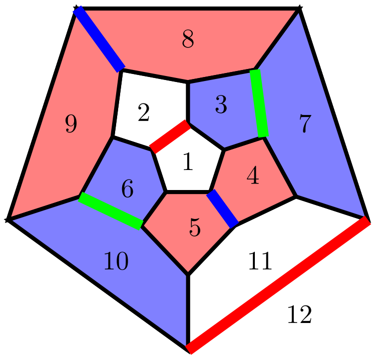


Figure 4: Color on $\mathbf{D}$ for the minimal common covering of $\mathbf{D}$ and $B_{4,4,4}$

On the other hand, we can color the dodecahedron with three colors as in Figure 4. The colors define the homomorphism

$$\begin{aligned} \rho : G_{\mathbf{D}} &\longrightarrow C_2 \times C_2 \times C_2 \\ x_1, x_2, x_{11}, x_{12} &\longmapsto (1,0,0) \\ x_4, x_5, x_8, x_9 &\longmapsto (0,1,0) \\ x_3, x_7, x_6, x_{10} &\longmapsto (0,0,1) \end{aligned} \tag{3.8}$$

and the monodromy

$$\begin{aligned} \omega_2 : G_{\mathbf{D}} &\longrightarrow \Sigma_8 \\ x_1, x_2, x_{11}, x_{12} &\longmapsto (1\,2)(3\,4)(5\,6)(7\,8) \\ x_4, x_5, x_8, x_9 &\longmapsto (1\,4)(2\,3)(5\,7)(6\,8) \\ x_3, x_7, x_6, x_{10} &\longmapsto (1\,6)(2\,5)(3\,6)(4\,8) \end{aligned}$$

defines the orbifold covering

$$p_2 : Z_{\mathbf{Q}} \longrightarrow \mathbf{D}$$

Observe that the three different colors on the three faces around a vertex implies that there are no singularities at the preimage of the vertex and the edges. The singular link corresponds to the preimage of the colored edges common to two faces with the same color. □

Remark 3.1 Let $\mathbf{Q}_2$ be the geometric orbifold structure on $\mathbf{Q}$ where all the angles are $\frac{\pi}{2}$, i.e., $\mathbf{Q}_2$ is the euclidean cube. The universal abelian cover $Z_{\mathbf{Q}_2}$ can be obtained also as a real moment-angle manifold, or an intersection of quadrics in $\mathbb{R}^6$, see Example 1.2, where we showed that it is homeomorphic to the three-torus $(\mathbb{S}^1)^3$ (see e.g. [LdM14]). Topologically $Z_{\mathbf{Q}} \equiv Z_{\mathbf{Q}_2}$ and this proves also that this topological manifold has the two following geometric structures: the euclidean one (with no singularities) and a hyperbolic one (with 12 circles with angle π).

As a consequence, the manifold $Z_{\mathbf{D}}$ is a 2^9-fold cover of the three-torus branched over a 12-component link.

Corollary 3.4 *The minimal common orbifold covering of the hyperbolic orbifolds* $\mathbf{D}$ *and* $B_{4,4,4}$ *which is a hyperbolic manifold is a two-fold covering of* $Z_{\mathbf{Q}}$*, made up with 16 dodecahedra.*

$$K \xrightarrow{2:1} Z_{\mathbf{Q}}$$

□

Corollary 3.5 *The Löbell manifold* $L(5)$ *is not an orbifold covering of the orbifold* $B_{4,4,4}$. □

Proposition 3.6 *The orbifold covering* $p_1 : Z_{\mathbf{Q}} \longrightarrow B_{4,4,4}$ *factors through a 4-fold orbifold covering* $p_3 : N \longrightarrow B_{4,4,4}$

Proof. The orbifold covering $p_3 : N \longrightarrow B_{4,4,4}$ is associated to the monodromy

$$\begin{aligned} \omega_3 : U &\longrightarrow \Sigma_4 \\ a &\longmapsto (1\,2)(3\,4) \\ b &\longmapsto (1\,3)(2\,4) \\ c &\longmapsto (1\,4)(2\,3) \end{aligned}$$

The cover N is a hyperbolic orbifold with a singular link of order 2. The fundamental group of the orbifold N is the kernel of ω_3, so it contains H_1 and has index 4 in U. The map p_3 is also a 4-fold locally cyclic covering of the sphere S^3 branched over the Borromean rings with branching index 2, or a 4-fold orbifold covering of $B_{2,2,2}$. The topological space N has both a Euclidean manifold structure (no singularity) and a hyperbolic orbifold structure with a singular link.

$$\begin{tikzcd} Z_{\mathbf{Q}} \arrow[rr, "2:1"] \arrow[dr, "2^3:1", "p_1"'] & & N \arrow[dl, "2^2:1"', "p_3"] \\ & B_{4,4,4} & \end{tikzcd}$$

□

Proposition 3.7 *The orbifold covering $p_2 : Z_{\mathbf{Q}} \longrightarrow \mathbf{D}$ factors through a 4-fold orbifold covering $p_4 : M \longrightarrow \mathbf{D}$*

Proof. The orbifold covering $p_4 : M \longrightarrow \mathbf{D}$ is associated to the monodromy

$$\begin{aligned} \omega_4 : G_{\mathbf{D}} &\longrightarrow \Sigma_4 \\ x_1, x_2, x_{11}, x_{12} &\longmapsto (1\,2)(3\,4) \\ x_4, x_5, x_8, x_9 &\longmapsto (1\,3)(2\,4) \\ x_3, x_7, x_6, x_{10} &\longmapsto (1\,4)(2\,3) \end{aligned}$$

The cover M is a hyperbolic orbifold with a singular link of order 2. The fundamental group of the orbifold M is the kernel of ω_4, so it contains H_1 and has index 4 in $G_{\mathbf{D}}$.

$$\begin{tikzcd} Z_{\mathbf{Q}} \arrow[rr, "2:1"] \arrow[dr, "2^3:1", "p_2"'] & & M \arrow[dl, "2^2:1"', "p_4"] \\ & \mathbf{D} & \end{tikzcd}$$

□

Proposition 3.8 *The 2-fold orbifold covering $p_5 : M_1 \longrightarrow M$ which is the 2-fold covering of M branched over the singular link, is the Löbell hyperbolic manifold $L(5)$.*

Proof. The hyperbolic orbifold cover is actually a hyperbolic manifold because all the branching index are 2 and the singular link in M have order 2. There are no singularities in M_1. Then $p_4 \circ p_5 : M_1 \longrightarrow \mathbf{D}$ is an 8-fold orbifold covering. It is a *small cover* of $\mathbf{D}$ ([DJ91]). It is proved in [GS03] that $\mathbf{D}$ has 25 small covers, but the only orientable one is the Löbell manifold $L(5)$. Then $M_1 = L(5)$. □

The following diagram summarizes all the relations among the studied orbifold coverings between the hyperbolic manifold $Z_{\mathbf{D}}$ and the orbifold $\mathbf{Q}$.

$$\begin{tikzcd} & & L(5) \arrow[r, "2"] & M \arrow[r, "2^2"] & \mathbf{D} \arrow[dr, "2^3"] & \\ Z_{\mathbf{D}} \arrow[r, "2^8"] & K \arrow[ur] \arrow[r, "2"] & Z_{\mathbf{Q}} \arrow[ur, "2"'] \arrow[dr, "2"] & & & \mathbf{Q} \\ & & & N \arrow[r, "2^2"'] & B_{4,4,4} \arrow[ur, "2^3"'] & \end{tikzcd} \tag{3.9}$$

This diagram has a counter-part for the fundamental orbifold groups. Here all the maps are inclusions.

$$
\begin{tikzcd}
 & & G_L \arrow[r, "2"] & \ker\omega_4 \arrow[r, "2^2"] & G_{\mathbf{D}} \arrow[dr, "2^3"] & \\
G'_{\mathbf{D}} \arrow[r, "2^8"] & G_K \arrow[r, "2"] \arrow[ur] & G'_{\mathbf{Q}} \arrow[ur, "2"'] \arrow[dr, "2"] & & & G_{\mathbf{Q}} \\
 & & & \ker\omega_3 \arrow[r, "2^2"'] & U \arrow[ur, "2^3"'] &
\end{tikzcd}
\tag{3.10}
$$

Theorem 3.9 *All the groups in (3.10) and* (3.5) *are commensurable subgroups of* $G_{\mathbf{Q}}$ *and they are arithmetic.*

Proof. The groups in (3.10) and (3.5) are part of the lattice of subgroups of the group $G_{\mathbf{Q}}$. All the inclusions in the diagram have finite index. Then all of them are commensurable subgroups. In [HLM92] the arithmeticity of the groups was studied for hyperbolic structures on the octant of the dodecahedron with different angles around the colored edges (Figure 3). One of then, $R(4,4,4)$, is the orbifold $\mathbf{Q}$. It is proved there that $\mathbf{Q} = R(4,4,4)$ is arithmetic. Therefore all the subgroups in (3.10) are arithmetic. In fact the concept of arithmetic subgroup is related with the problem of enumerating all the *forms of Clifford-Klein* or *geometric orbifolds of constant curvature, complete and with finite volume*, in actual language, as was pointed out in the historical comments contained in [Mon13]. □

It is possible to compute all the groups in (3.10) but most of them have very long presentations. For example, $\pi_1(Z_{\mathbf{D}}) = G'_{\mathbf{D}}$ can be computed using `Sagemath` [S+15] and `GAP4` [GAP15]. We obtain a group whose abelianization is a free abelian group of rank 935. It is possible to find a presentation with 935 generators and 955 relations (all of them product of commutators).

References

[BBCG10] A. Bahri, M. Bendersky, F. R. Cohen, and S. Gitler: *The polyhedral product functor: a method of decomposition for moment-angle complexes, arrangements and related spaces*. Adv. Math. **225** (2010), no. 3, 1634–1668.

[DJ91] M. W. Davis, T. Januszkiewicz: *Convex polytopes, Coxeter orbifolds and torus actions*. Duke Math. J. **62** (1991), no. 2, 417–451.

[GAP15] The GAP Group: *GAP – Groups, Algorithms, and Programming, Version 4.7.7.* 2015.

[GLdM13] S. Gitler, S. López de Medrano: *Intersections of quadrics, moment-angle manifolds and connected sums*. Geom. Topol. **17** (2013), no. 3, 1497–1534.

[GS03] A. Garrison, R. Scott: *Small covers of the dodecahedron and the 120-cell*. Proc. Amer. Math. Soc. **131** (2003), no. 3, 963–971.

[HLM92] H. M. Hilden, M. T. Lozano, J. M. Montesinos: *On the Borromean orbifolds: geometry and arithmetic*. Topology '90 (Columbus, OH, 1990), Ohio State Univ. Math. Res. Inst. Publ., vol. 1, de Gruyter, Berlin, 1992, pp. 133–167.

[HLMW87] H. M. Hilden, M. T. Lozano, J. M. Montesinos, W.C. Whitten: *On universal groups and three-manifolds*. Invent. Math. **87** (1987), no. 3, 441–456.

[LdM89] S. López de Medrano: *Topology of the intersection of quadrics in* $\mathbf{R}^n$. Algebraic topology (Arcata, CA, 1986), Lecture Notes in Math., vol. 1370, Springer, Berlin, 1989, pp. 280–292.

[LdM14] ———: *Singularities of homogeneous quadratic mappings*. Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Math. RACSAM **108** (2014), no. 1, 95–112.

[Löb31] F. Löbell: *Beispiele geschlossener dreidimensionaler Clifford-Kleinscher räume negativer krümmung*. Berichte Leipzig **83** (1931), 167–174.

[Mon13] J. M. Montesinos: *On integral quadratic forms having commensurable groups of automorphisms*. Hiroshima Math. J. **43** (2013), no. 3, 371–411.

[S$^+$15] W. A. Stein et al.: *Sage Mathematics Software (Version 6.6)*. The Sage Development Team, 2015, `http://www.sagemath.org`.

[Thu80] W. P. Thurston: *The geometry and topology of three-manifolds*. Princeton University, available at `http://library.msri.org/nonmsri/gt3m`, 1980.

[Ves98] A. Yu. Vesnin: *Volumes of Löbell* 3*-manifolds*. Mat. Zametki **64** (1998), no. 1, 17–23.

[Ves10] ———: *Volumes and normalized volumes of right-angled hyperbolic polyhedra*. Atti Semin. Mat. Fis. Univ. Modena Reggio Emilia **57** (2010), 159–169 (2011).

[Wal80] C. T. C. Wall: *Stability, pencils and polytopes*. Bull. London Math. Soc. **12** (1980), no. 6, 401–421.

ρ-pairs in graphs representing surfaces

Paola Bandieri *

Dipartimento di Scienze Fisiche, Informatiche e Matematiche
Università di Modena e Reggio Emilia
Via Campi 213 B
I-41125 MODENA (Italy)
paola.bandieri@unimore.it

Dedicated to Professor Josè Maria Montesinos-Amilibia on the occasion of his 70th birthday.

Abstract

In this work, we study the effects of the switchings of ρ-pairs in 3-colored graphs representing surfaces, possibly with (connected or not) boundary.

2000 Mathematics Subject Classification: Primary 57Q15; Secondary 57M15.
Keywords: colored graphs, ρ-pairs, compact surfaces.

1. Introduction

Crystallization theory – arising from an idea of M. Pezzana – makes use of a particular class of edge-colored graphs to represent PL-manifolds. It is a helpful combinatorial tool for the study of PL-manifolds and it is strictly related, in low dimensions, with other representation theories, some classical, such as Heegaard diagrams ([18]), some other more recent, such as special spines ([21]) and face-pairing graphs ([6]).

However, the main advantage of crystallization theory is that it is easy to extend it to any dimension n and to all compact PL n-manifolds, without restrictions on orientability, boundary, connectedness, etc. (see [3] and [12] for a survey about the theory); moreover, crystallizations are suitable for computer manipulation in order to generate and classify catalogues of triangulations of manifolds ([4]).

A classical problem, common to all representation theories, is that of defining transformations (*moves*) on the objects changing one (for example, an edge-colored graph) into another equivalent object (in our context, another edge-colored graph

*Work supported by the "National Group for Algebraic and Geometric Structures, and their Applications" (GNSAGA - INDAM) and by M.I.U.R. of Italy (project "Strutture Geometriche, Combinatoria e loro Applicazioni").

representing the same PL-manifold). A key concept to define a useful equivalence relation between edge-colored graphs is that of color-isomorphism, described in [12].

More recently, Lins, in [20], introduced the concept of *ρ-pair*, defining a particular subgraph in 4-colored graphs representing orientable, closed 3-manifolds and a related move (*switching of ρ-pairs*); subsequently, he and Ferri extended this concept to 4-colored graphs representing non-orientable 3-manifolds ([13]). In [5], the definition of ρ-pair is further extended to dimension n, i.e. to $(n+1)$-colored graphs representing closed n-manifolds.

We point out that the graph obtained by the switching of a ρ-pair represents, in general, a different manifold from the one represented by the starting graph; in the quoted papers the effects of switching ρ-pairs are described and the relations are obtained between a colored graph and the colored graph obtained from it by the switching.

Hence, this move allows to recognize possible equivalent graphs since its effects, on the represented manifolds, are completely determined. As a consequence, the switchings of ρ- pairs, in particular for 4- and 5-colored graphs turned out to be very useful in order to generate and classify efficient catalogues of 3-manifolds (see [1], [2] and [8]) and of 4-manifolds (see [9], [10] and [7]).

In this paper, we introduce the concept of ρ-pair for 3-colored graphs (possibly with boundary), representing connected surfaces, possibly with (connected or not) boundary; finally, we analyze the effects of switching ρ-pairs, the relations existing between the starting graph and the obtained one and, consequently, between the represented surfaces.

2. Notations

In the following, all manifolds will be compact and, when not otherwise stated, connected. For the basic notions of PL topology, we refer to [24] and to [16]; with the symbol "$\simeq$", we will mean *PL-homeomorphic*; with the symbol "$\#$", we will mean *connected sum* and, finally, with the symbol "$\#_\partial$", we will mean *connected sum along a boundary component.*

For graph theory, see [17] and [25].

The term "graph" will be used instead of "multigraph", hence multiple edges are allowed, but loops are forbidden. $V(\Gamma)$ and $E(\Gamma)$ will denote the vertex-set and the edge-set of the graph Γ.

A *disconnecting edge-set* or *edge-cut* [11] in a connected graph Γ is a set Ψ of edges of Γ such that $\Gamma \setminus \Psi$ is disconnected. In particular, if $\Psi = \{\mathbf{e}\}$, then $\mathbf{e}$ is called a *bridge.*

In this paper, a *path* (resp. a *cycle*) is a finite sequence of vertices and edges $\{v_1, e_1, \dots, v_k, e_k, v_{k+1}\}$ such that $v_j, v_{j+1}, j = 1, \dots, k$ are the ends of e_j and $v_h \neq v_l, h \neq l$ (resp. $v_h \neq v_l, h, l \neq 1, k+1$, if $h \neq l$, and $v_1 = v_{k+1}$).

An $(n+1)$-*colored graph* is a pair (Γ, γ), where Γ is a graph and $\gamma : E(\Gamma) \to \Delta_n = \{0, 1, \dots, n\}$ is a proper coloration of the edges of Γ, that is, a map injective on the star of each vertex $v \in V(\Gamma)$. Usually, we will denote simply by Γ the graph (Γ, γ).

As a consequence of the previous definition, any vertex $v \in V(\Gamma)$ has degree $\leqslant n$; we will call v an *internal vertex*, if its degree is exactly n, *boundary vertex*, otherwise. Obviously, if Γ is regular of degree n, all its vertices are internal.

From now on, we will call Γ *without boundary* if it is regular of degree n and *with boundary*, otherwise. When we state a property for an "$(n+1)$-*colored graph* Γ", if not otherwise specified, it means that such a property holds both for graphs without and for graph with boundary.

Let B be a subset of Δ_n. Then, we denote by Γ_B the subgraph $(V(\Gamma), \gamma^{-1}(B))$. The connected components of Γ_B will be called B-*residues* of (Γ, γ); moreover, g_B denotes the number of B-residues of Γ. For each $i \in \Delta_n$, we set $\hat{\imath} = \Delta_n \setminus \{i\}$ and $g_{\hat{\imath}}$ the number of components of the graph $\Gamma_{\hat{\imath}}$, obtained by deleting all edges colored i from Γ. Moreover, if $B = \{i, j\}$, we write g_{ij} instead of $g_{\{i,j\}}$. Note that, even in the case of Γ being non-bipartite, the $\hat{\imath}$-residues of Γ are bipartite, for each color i. Moreover, for each graph Ψ, $g(\Psi)$ will denote the number of connected components of Ψ. In particular, if Ψ is the graph having the empty-set both as vertex-set and as edge-set ($\Psi = \emptyset$), we set, by convention, $g(\Psi) = 1$.

An $(n+1)$-colored graph (Γ, γ) is said to be *regular with respect to the color* n, if $\Gamma_{\hat{n}}$ is a regular graph of degree n. Hence, all boundary vertices, if any, of Γ have degree n and the missing color is always color n. In the following, all $(n+1)$-colored graphs will be regular with respect to the color n.

For each $(n+1)$-colored graph (Γ, γ), we can define a (possibly non-connected) *boundary graph* $(\partial\Gamma, {}^{\partial}\gamma)$ as follows:

- $V(\partial\Gamma)$ is the set of boundary vertices of Γ;
- two vertices $\mathbf{v}, \mathbf{w}$ of $V(\partial\Gamma)$ are i-adjacent, $i = 0, 1, \dots, n-1$, in $\partial\Gamma$ iff there exists an (i, n)-colored path in Γ joining $\mathbf{v}$ and $\mathbf{w}$.

Obviously, if Γ is regular of degree n, i.e. all its vertices are internal, then $(\partial\Gamma, {}^{\partial}\gamma)$ is the empty graph, otherwise it is a regular graph of degree n. Moreover, ${}^{\partial}\gamma$ is a proper coloration of the edges of $\partial\Gamma$ and $\partial(\partial\Gamma) = \emptyset$.

We can construct the pseudocomplex ([19]) $K = K(\Gamma)$, associated to Γ, as follows:

(a) for each vertex $\mathbf{v} \in V(\Gamma)$, choose an n-simplex $\sigma(\mathbf{v})$ and label by Δ_n its vertices (0-dimensional faces);

(b) if $\mathbf{v}, \mathbf{u}$ are c-adjacent vertices of Γ, $c \in \Delta_n$, then identify the equally colored edges of $\sigma(\mathbf{v})$ and $\sigma(\mathbf{u})$ which are opposite to the c-labelled vertices of $\sigma(\mathbf{v})$ and of $\sigma(\mathbf{u})$.

Γ is said *to represent* both $K(\Gamma)$ and $|K(\Gamma)|$.

Note that the connected components of $(\partial\Gamma, {}^{\partial}\gamma)$ represent the boundary components of $|K(\Gamma)|$.

$|K(\Gamma)|$ is orientable iff Γ is bipartite ([12]).

A *colored n-complex* is a pseudocomplex K of dimension n, with a labelling of its vertices by Δ_n which is injective on the vertex-set of each n-simplex of K.

Note that $K(\Gamma)$ inherits from γ a natural vertex-coloration ζ, which is injective on each n-simplex of $K(\Gamma)$. Moreover, for each $i \in \Delta_n$, the number of i-colored vertices of $K(\Gamma)$ is $g_{\hat{\imath}}$. Both Γ and $K(\Gamma)$ are called *contracted* iff $g_{\hat{\imath}} = 1$, for $i = n$ and $g_{\hat{\imath}} = g(\partial\Gamma)$, for $i \neq n$. If $|K(\Gamma)|$ is a manifold and the graph Γ is contracted, then it is called a *crystallization* of $|K(\Gamma)|$.

Note that, if Γ is a graph without boundary, then $\partial\Gamma = \emptyset$ and $g_{\hat{\imath}} = 1$, for $i \in \Delta_n$.

It is well known that every PL-manifold admits crystallizations (see [22] and [23]).

In the present paper, we are interested in surfaces (2-manifolds), hence, in the following, we set $n = 2$.

We will denote by $T_{g,\lambda}$ (resp. $N_{h,\lambda}$) the orientable (resp. non-orientable) surface of genus g (resp. h), with λ boundary components (or *holes*). For short, we will denote by T_g (resp. N_h) the closed surface $T_{g,0}$ (resp. $N_{h,0}$) and by $\mathbb{H}$ we will denote either T_1 or N_2, whereas, by $\mathbb{H}_1$ we will denote either $T_{1,1}$ or $N_{2,1}$.

If Γ is a 3 - colored graph, then $|K(\Gamma)|$ is a (closed or not) surface. In this case, for each $i \in \Delta_2$, the number $g_{\hat{\imath}}$ of components of $\Gamma_{\hat{\imath}}$ is the sum of the number of its bicolored cycles and the number of its bicolored paths.

Note that setting $\Delta_2 = \{c, d, k\}$, with the above notations, we have $g_{\hat{c}} = g_{dk}$; moreover, in the following, by $\dot{g}_{dk} = \dot{g}_{dk}(\Gamma)$ we will denote the number of $\{d, k\}$-colored cycles of Γ (i.e. the number of cycles with edges alternatively colored d, k).

As it is well known, for each 3-colored graph (Γ, γ), there is a unique surface $|K(\Gamma)|$ associated to Γ. Because of the above considerations, $|K(\Gamma)|$ is orientable iff Γ is bipartite, and the number of its boundary components is equal to the number of connected components (*holes*) of $\partial\Gamma$, hence $|K(\Gamma)|$ is closed iff Γ is without boundary. For sake of conciseness, in the following, we will often call boundary components of Γ the connected components of $\partial\Gamma$.

Note that, by construction, the number of 2-simplexes of $K(\Gamma)$ coincides with the number of vertices of Γ, the number of 1-simplexes of $|K(\Gamma)|$ coincides with the number of edges of Γ and the number of 0-simplexes of $|K(\Gamma)|$ coincides with the number of bicolored cycles (or path) of Γ, then, for each graph Ψ representing the surface $|K(\Gamma)|$ the Euler characteristic of Ψ coincides with the Euler Characteristic of $|K(\Gamma)|$. Hence, in the following, χ will denote both the Euler characteristic of $|K(\Gamma)|$ and of Γ (and of any graph Ψ representing $|K(\Gamma)|$), that is $\chi = \chi(|K(\Gamma)|) = \chi(\Gamma)$ and the calculation can be made directly on Γ, as follows: $\chi(|K(\Gamma)|) = \dot{g}_{cd} + \dot{g}_{ck} + \dot{g}_{dk} - \frac{\dot{p}}{2}$, where, of course, if Γ is without boundary, all connected components of $\Gamma_{\{c,d\}}$ are cycles, hence $g_{dk} = \dot{g}_{dk}$ and $\dot{p} = p$.

Hence, if (Γ, γ) is a bipartite (resp. non-bipartite) 3-colored graph without boundary, it represents the closed, orientable (resp. non-orientable) surface of genus $g = 1 - \frac{\chi}{2} = 1 - \frac{1}{2}g_{cd} - \frac{1}{2}g_{ck} - \frac{1}{2}g_{dk} + \frac{p}{4}$ (resp. $h = 2 - \chi = 2 - g_{cd} - g_{ck} - g_{dk} + \frac{p}{2}$). In particular, if Γ is a bipartite (resp. non-bipartite) crystallization of a closed surface, then it represents the orientable (resp. non-orientable) surface T_g (resp. N_h) of genus $g = 1 - \frac{\chi}{2} = \frac{p}{4} - \frac{1}{2}$ (resp. $h = 2 - \chi = \frac{p}{2} - 1$).

If (Γ, γ) is a bipartite (resp. non-bipartite) 3-colored graph with λ boundary components, it represents the closed, orientable (resp. non-orientable) surface of genus $g = 1 - \frac{\chi}{2} - \frac{\lambda}{2}$, if Γ is bipartite and $h = 2 - \chi - \lambda$, if Γ is non-bipartite.

In the following, for each graph Ψ and for each $\mathbf{x}$ either in $E(\Psi)$ or in $V(\Psi)$, we denote by $\Psi(\mathbf{x})$ the connected component of Ψ containing $\mathbf{x}$. Moreover, for each $\mathbf{x}$ either in $E(\Psi)$ or in $V(\Psi)$, we simply will write $\mathbf{x} \in \Psi$.

3. ρ-pairs and ρ^*-pairs in 3-colored graphs

If Γ is an $(n+1)$ - colored graph without boundary, a ρ_m - pair (for short ρ-pair) in Γ is a pair $R = \{e, f\}$ of edge of the same color - say c - such that e, f lie on the same $\{c, i_h\}$ - colored cycles, for m different colors $i_1, \ldots, i_m \in \Delta_n$ and on different $\{c, d\}$ - colored cycles for $d \neq i_1, \ldots, i_m$. In general, they are interesting in the cases $m = n - 1$ and $m = n$ (see [20], for $n = 3$, and [5], for $n \geq 3$).

In the following, we analyse the concept of ρ - pair for surfaces and extend it to the boundary case.

Definition 3.1 *Let Γ be a 3-colored graph and let $\mathbf{u}, \mathbf{v}$ be boundary vertices in Γ. The pair $\{\mathbf{u}, \mathbf{v}\}$ is a wound of type 2 or simply a 2-wound involving colors 0 and 1 iff $\Gamma_{\{j,2\}}(\mathbf{u}) = \Gamma_{\{j,2\}}(\mathbf{v})$, with $j = 0, 1$.*

In the following, for short, we will call simply *wound* a 2-wound.

If $\{\mathbf{u}, \mathbf{v}\}$ is a wound, then, by adding a 2-colored edge between $\mathbf{u}$ and $\mathbf{v}$, we obtain a new graph $\bar{\Gamma}$ and call it the graph obtained from Γ by *suturing the wound*. In the following, we will denote by $\mathbf{f}$ the 2-colored edge between $\mathbf{u}$ and $\mathbf{v}$ in $\bar{\Gamma}$, and call it the *suture* of $\{\mathbf{u}, \mathbf{v}\}$.

Of course, $\bar{\Gamma}$ has two internal vertices more than Γ and the same number of $\{0,1\}$-colored cycles. Furthermore, $\bar{\Gamma}$ has one $\{0,2\}$-colored cycle and one $\{1,2\}$-colored cycle more. Hence, the Euler characteristic of the associated surface $|K(\bar{\Gamma})|$ coincides with $\chi + 1 = \chi|K(\Gamma)| + 1$.

From the above definitions, considerations and the results of [14], we deduce:

Proposition 3.1 *If $\{\mathbf{u}, \mathbf{v}\}$ is a 2-wound, its suture produces a surface obtained from $|K(\Gamma)|$ by capping off a boundary component with a 2-disc, that is*

$$|K(\bar{\Gamma})| \simeq |K(\Gamma)| \#_\partial T_{0,1}.$$

Definition 3.2 *Let $R = (\mathbf{e}, \mathbf{f})$ be a pair of edges of Γ. R is a ρ_1-pair of color $c \in \Delta_2$, not involving color d, iff we have:*

- $\gamma(\mathbf{e}) = \gamma(\mathbf{f}) = c$;
- $\Gamma_{ck}(\mathbf{e}) = \Gamma_{ck}(\mathbf{f})$;
- $\Gamma_{cd}(\mathbf{e}) \neq \Gamma_{cd}(\mathbf{f})$.

Definition 3.3 *Let $R^* = (\mathbf{e}, \{\mathbf{u}, \mathbf{v}\})$ be a pair, where $\mathbf{e}$ is an edge of Γ and $\{\mathbf{u}, \mathbf{v}\}$ is a wound in Γ. R^* is a ρ_1^*-pair, not involving color d, in Γ iff $R = (\mathbf{e}, \mathbf{f})$ is a ρ_1-pair of color 2, not involving color d, in $\bar{\Gamma}$.*

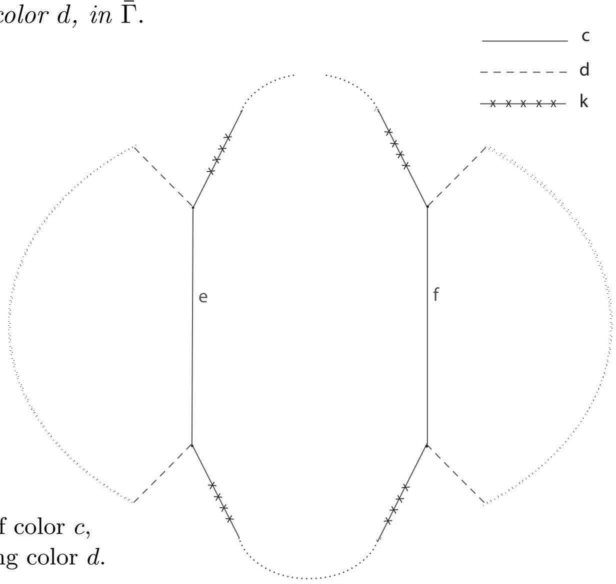


Figure 1a: A ρ_1-pair of color c, not involving color d.

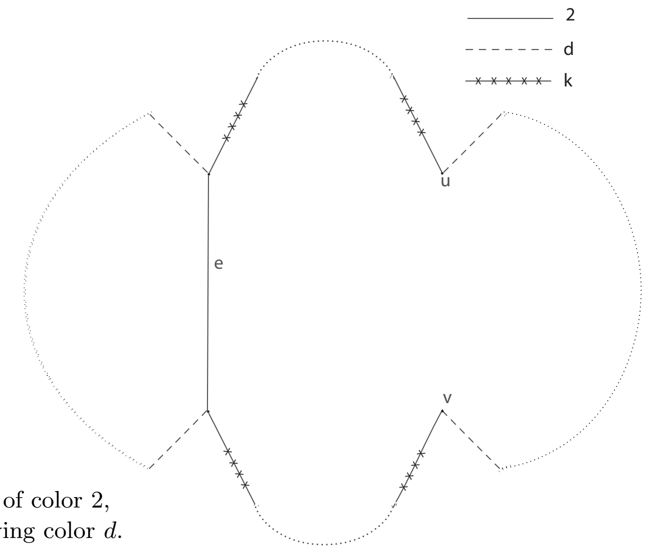


Figure 1b: A ρ_1^*-pair of color 2, not involving color d.

In the following, for a ρ_1 - (resp. a ρ_1^* -) pair, when not otherwise specified, c (resp. 2) will denote the color of the pair and d the not involved one.

Definition 3.4 *Let $R = (\mathbf{e}, \mathbf{f})$ be a pair as above. R is a ρ_2-pair of color $c \in \Delta_2$ iff we have:*

- *$\gamma(\mathbf{e}) = \gamma(\mathbf{f}) = c$;*
- *$\Gamma_{ci}(\mathbf{e}) = \Gamma_{ci}(\mathbf{f})$, for each color $i \in \Delta_2 \setminus \{c\}$.*

Definition 3.5 *Let $R^* = (\mathbf{e}, \{\mathbf{u}, \mathbf{v}\})$ be a pair as above. R^* is a ρ_2^*-pair of color 2 in Γ iff $R = (\mathbf{e}, \mathbf{f})$ is a ρ_2-pair in $\bar{\Gamma}$.*

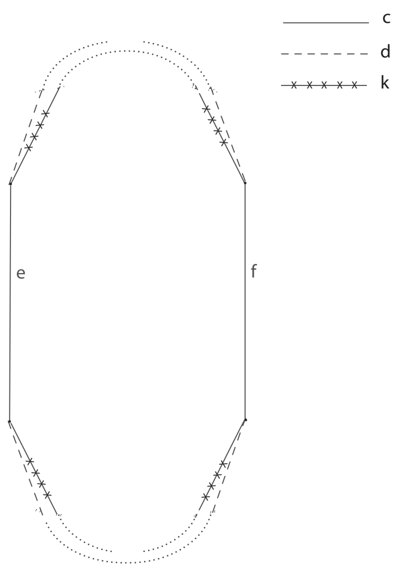


Figure 2a: A ρ_2-pair of color c.

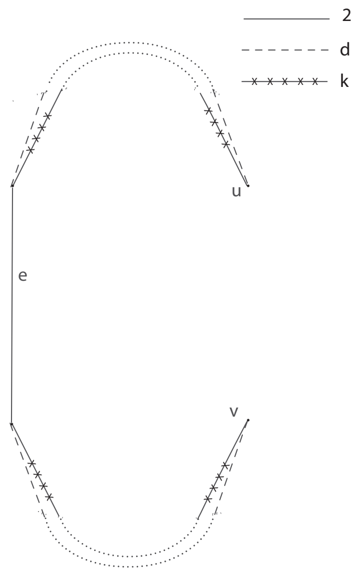


Figure 2b: A ρ_2^*-pair of color 2.

Note that, for $h = 1, 2$, any ρ_h^*-pair in Γ_h corresponds to a ρ_h-pair of color 2 in $\bar{\Gamma}$, so from now on we will write "*R is a ρ_h-pair in* $\bar{\Gamma}$", to mean "*R^* is a ρ_h^*-pair in* Γ".

In the following, we will also call *ρ-pair*, for short, either a ρ_1-pair or a ρ_2-pair, both in Γ and in $\bar{\Gamma}$.

Remark 3.1 *If R is a ρ-pair in $\bar{\Gamma}$, then it is a wound.*

Given a ρ-pair $R = (\mathbf{e}, \mathbf{f})$, fix arbitrarily an orientation of $\mathbf{e}$ and set $\mathbf{e_0}$ the first end (*tail*) of $\mathbf{e}$ and $\mathbf{e_1}$ the second end (*head*) of $\mathbf{e}$, according with the fixed orientation.

Note that for each 3-colored graph Γ and for each color i in Δ_2, since $\Gamma_{\hat{\imath}}$ is always bipartite, it induces a partition $V = V_0^i \cup V_1^i$ of the vertex set of Γ. For each color $i \in \Delta_2$, we choose the above partition so that $\mathbf{e_0} \in V_0^i$ and $\mathbf{e_1} \in V_1^i$. The arbitrarily fixed orientation of $\mathbf{e}$ induces an orientation on $\Gamma_{\hat{\imath}}(\mathbf{e})$ (resp. on $\bar{\Gamma}_{\hat{\imath}}(\mathbf{e})$), for each $i \in \Delta_2$.

Given a 3-colored graph Γ and a ρ_1-pair R in Γ, we consider the bipartition induced by $\Gamma_{\hat{d}}$ to label the ends of $\mathbf{f}$, following the cycle $\Gamma_{ck}(\mathbf{e})$, according to the arbitrarily fixed orientation of $\mathbf{e}$.

If R is a ρ_2-pair of Γ (resp. of $\bar{\Gamma}$), being $c \in \Delta_2$ (resp. $c = 2$) the color involved in the ρ_2-pair R and, as above, $\Delta_2 \setminus \{c\} = \{d, k\}$, then set $\mathbf{f_0^i}$ the tail of $\mathbf{f}$ and $\mathbf{f_1^i}$ its head in $\Gamma_{\hat{\imath}}(\mathbf{e})$ (resp. in $\bar{\Gamma}_{\hat{\imath}}(\mathbf{e})$), $i = d, k$.

We can distinguish two cases:

(C_1) $\mathbf{f_0^k} = \mathbf{f_0^d}$ and $\mathbf{f_1^k} = \mathbf{f_1^d}$;

(C_2) $\mathbf{f_0^k} = \mathbf{f_1^d}$ and $\mathbf{f_1^k} = \mathbf{f_0^d}$.

Remark 3.2 *If Γ is bipartite, we can have only case (C_1); otherwise, both cases (C_1) and (C_2) can occur.*

Definition 3.6 *Fix a color $i \in \Delta_2 \setminus c$. A ρ-pair $R = (\mathbf{e}, \mathbf{f})$, involving color c in Γ (resp. in $\bar{\Gamma}$), is called an edge-cut with respect to color i (resp. $\mathbf{e}$ is a bridge with respect to color i) if it is an edge-cut in Γ (resp. in $\bar{\Gamma}$) such that $\mathbf{e_0}$ and $\mathbf{f_1^i}$ lie in the same connected component of $\Gamma \setminus \{\mathbf{e}, \mathbf{f}\}$ (resp. of $\bar{\Gamma} \setminus \{\mathbf{e}, \mathbf{f}\}$) and $\mathbf{e_1}$ and $\mathbf{f_0^i}$ lie in the other connected component of $\Gamma \setminus \{\mathbf{e}, \mathbf{f}\}$ (resp. of $\bar{\Gamma} \setminus \{\mathbf{e}, \mathbf{f}\}$).*

Definition 3.7 *With the above notations, we call switching of the ρ_1-pair R of color c in Γ or in $\bar{\Gamma}$, the following move:*

- *delete $\mathbf{e}, \mathbf{f}$ from Γ;*
- *introduce edges joining $\mathbf{e_0}$ with $\mathbf{f_1^d}$ and $\mathbf{e_1}$ with $\mathbf{f_0^d}$.*

Figure 3 shows locally a switching of a ρ_1-pair, both in Γ and in $\bar{\Gamma}$.

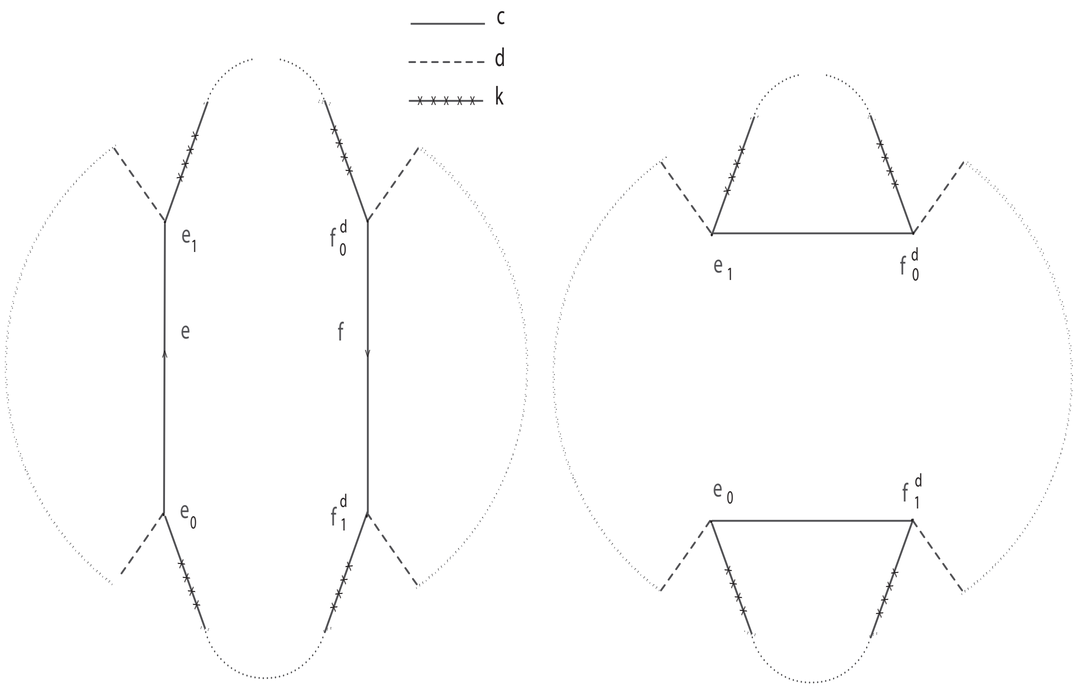


Figure 3: Switching of a ρ_1-pair.

If Γ is a 3-colored graph and $R = (\mathbf{e}, \mathbf{f})$ is a ρ_1-pair in Γ (resp. in $\bar{\Gamma}$), then we denote by Γ' the 3-colored graph obtained from Γ by performing the switching of R and we say that Γ' is obtained from Γ by the switching of R.

Note that the above move always preserves the bipartition of $\Gamma_{\hat{d}}$, and the bipartition of Γ, in the case of Γ being bipartite.

Definition 3.8 *With the above notations, we call switching of a ρ_2-pair $R = (\mathbf{e}, \mathbf{f})$ of Γ or of $\bar{\Gamma}$ with respect to the color $i \in \{d, k\}$ the following move:*

- *delete $\mathbf{e}, \mathbf{f}$ from Γ;*
- *introduce edges joining $\mathbf{e_0}$ with $\mathbf{f_1^i}$ and $\mathbf{e_1}$ with $\mathbf{f_0^i}$.*

Remark 3.3 *If R is a ρ_2-pair of Γ or $\bar{\Gamma}$, then by performing the moves of Definitions 7 and 8, there are two possibly different resulting graphs: (Γ'_k, γ'_k), which is obtained from Γ by switching with respect to color k and (Γ'_d, γ'_d), which is obtained from Γ by switching with respect to color d.*

Obviously, in case (C_1), we have $(\Gamma'_k, \gamma'_k) = (\Gamma'_d, \gamma'_d) = (\Gamma', \gamma')$.

Figures 4a, 4b and 4c present locally the effects of switching a ρ_2-pair, both in Γ and in $\bar{\Gamma}$. Figure 4a shows case (C_1) and Figure 4b and Figure 4c show case (C_2).

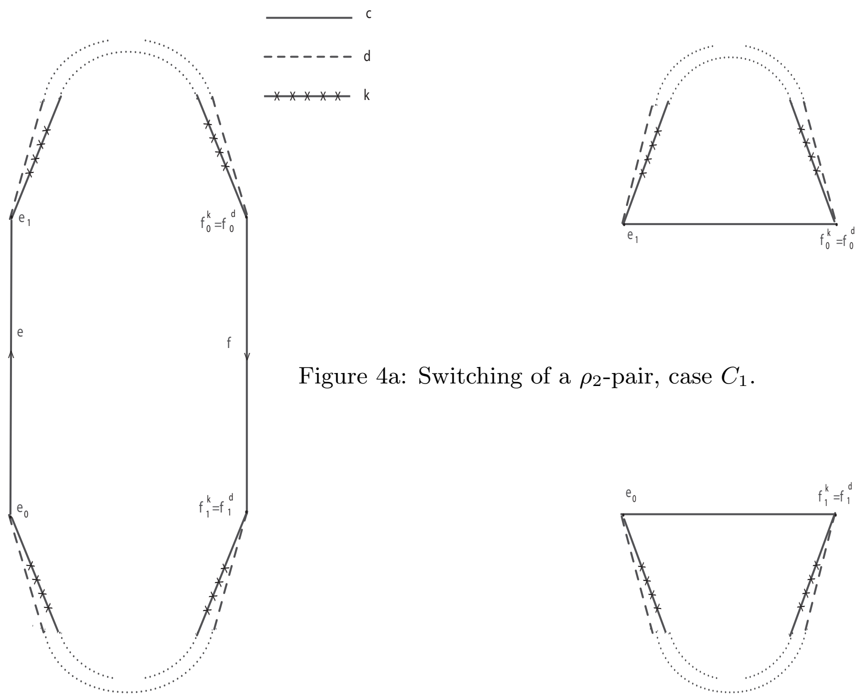


Figure 4a: Switching of a ρ_2-pair, case C_1.

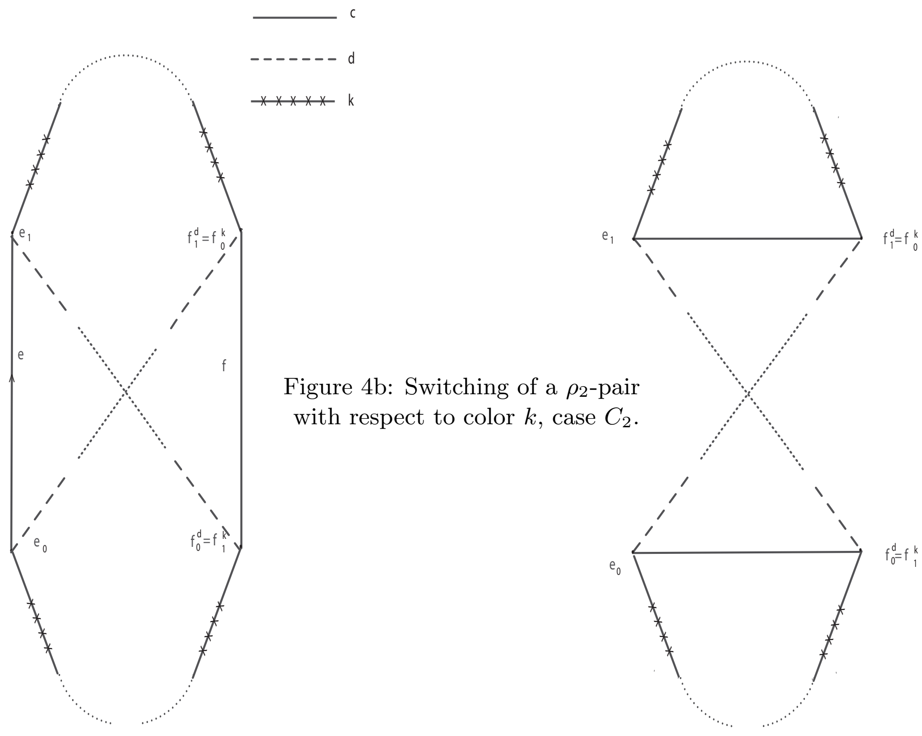


Figure 4b: Switching of a ρ_2-pair
with respect to color k, case C_2.

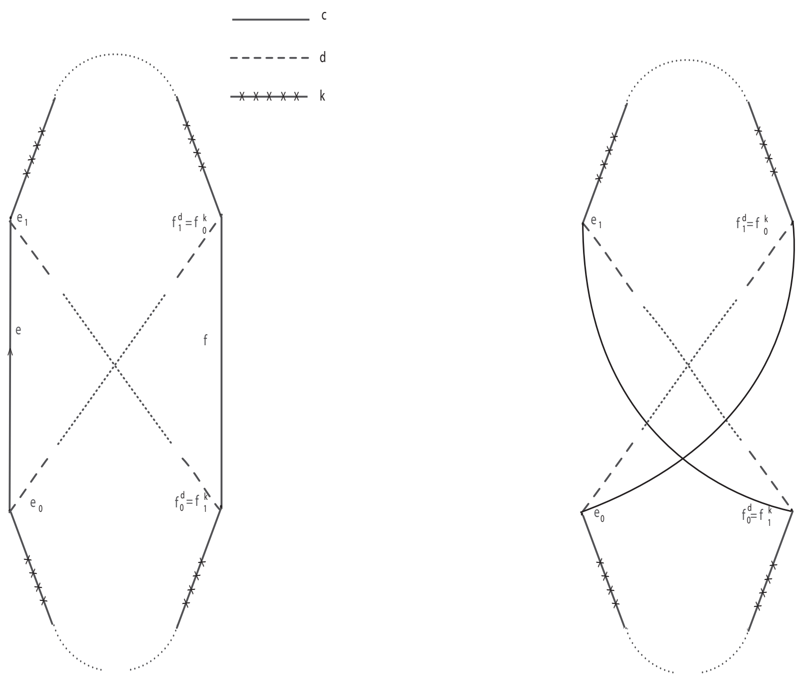


Figure 4c: Switching of a ρ_2-pair with respect to color d, case C_2.

Theorem 3.2 *Let Γ be a 3-colored graph, let $R = (\mathbf{e}, \mathbf{f})$ be a ρ-pair of color c of Γ (resp. of $\bar{\Gamma}$) and let Γ' be the graph obtained from Γ by any switching of R. Then, either Γ' is connected or Γ' has two connected components.*

Proof. We distinguish the cases of Γ without or with boundary.

Let R be a ρ-pair of color c in Γ (resp. of color 2 in $\bar{\Gamma}$). Denote, as above, by $\mathbf{e}_0, \mathbf{e}_1, \mathbf{f}_0^k, \mathbf{f}_1^k, \mathbf{f}_0^d, \mathbf{f}_1^d$ the ends of the edges $\mathbf{e}, \mathbf{f}$ in Γ (resp. in $\bar{\Gamma}$).

STEP 1: Γ *without boundary.*

Let $R = (\mathbf{e}, \mathbf{f})$ be a ρ_1-pair in Γ. Then, the graph $\Gamma \setminus \{\mathbf{e}, \mathbf{f}\}$ is connected, because $\Gamma_{cd}(\mathbf{e}) \neq \Gamma_{cd}(\mathbf{f})$. Hence, the graph Γ', obtained by the switching of $R = (\mathbf{e}, \mathbf{f})$ with respect to the color d in Γ, is connected, too.

Let now $R = (\mathbf{e}, \mathbf{f})$ be a ρ_2-pair; then, obviously, either the graph $\Gamma \setminus \{\mathbf{e}, \mathbf{f}\}$ is connected, or R is an edge-cut and $\Gamma \setminus \{\mathbf{e}, \mathbf{f}\}$ has two connected components. Hence, after the switching of R, either Γ' is connected or it has two connected components, too.

STEP 2: Γ *with boundary.*

If $R = (\mathbf{e}, \mathbf{f})$ is a ρ_1-pair of Γ (resp. of $\bar{\Gamma}$) and an edge-cut with respect to color d, then the graph $\Gamma \setminus \{\mathbf{e}, \mathbf{f}\}$ splits into two connected components, hence the graph Γ', obtained by the switching of $R = (\mathbf{e}, \mathbf{f})$ with respect to the color d in Γ (resp. in $\bar{\Gamma}$) , splits into two connected components, too; on the other hand, if the pair $R = (\mathbf{e}, \mathbf{f})$ is a ρ_1-pair of Γ (resp. of $\bar{\Gamma}$), but it is not an edge-cut with respect to the color d, then there is in $\Gamma \setminus \{\mathbf{e}, \mathbf{f}\}$ at least one path joining the ends of $\mathbf{e}$ or those of $\mathbf{f}$; therefore the graph Γ', obtained by the switching of $R = (\mathbf{e}, \mathbf{f})$ in Γ (resp. in $\bar{\Gamma}$), is connected.

If $R = (\mathbf{e}, \mathbf{f})$ is a ρ_2-pair, let us consider case (C_1). If R is an edge-cut with respect both to color d and to color k, then, since there is a unique way to perform the switching of R, the graph Γ', splits into two connected components.

If the ρ_2-pair $R = (\mathbf{e}, \mathbf{f})$ is not an edge-cut with respect to one of the colors, say d, then there is a (c, d)-colored path joining the ends of $\mathbf{e}$ or those of $\mathbf{f}$. Therefore, $\Gamma \setminus \{\mathbf{e}, \mathbf{f}\}$ is connected and Γ' is connected, too.

With regard to case (C_2), note that R cannot be an edge-cut with respect both to d and to k, hence Γ' is always connected. $\square$

Remark 3.4 *Note that, if Γ is a graph without boundary, Γ' is without boundary, too.*

Remark 3.5 *Let Γ be a 3-colored graph and let R be a ρ-pair either in Γ or in $\bar{\Gamma}$. Let Γ' denote the graph obtained by the switching of R.*

- *If R is a ρ_1-pair of color c, not involving color d in Γ, and if at least one of $\Gamma_{cd}(\mathbf{e})$ or $\Gamma_{cd}(\mathbf{f})$ is a cycle, then (Figure 3) by the switching of R in Γ' it has $\dot{g}'_{cd} = \dot{g}_{cd} - 1$, $\dot{g}'_{ck} = \dot{g}_{ck} + 1$ and, finally, $\dot{g}'_{dk} = \dot{g}_{dk}$;*
- *if R is a ρ_1-pair of color c, not involving color d in Γ, and if both $\Gamma_{cd}(\mathbf{e})$ and $\Gamma_{cd}(\mathbf{f})$ are paths, then by the switching of R in Γ' it has $\dot{g}'_{cd} = \dot{g}_{cd}$, $\dot{g}'_{ck} = \dot{g}_{ck} + 1$ and, finally, $\dot{g}'_{dk} = \dot{g}_{dk}$;*
- *if R is a ρ_2-pair of color c in Γ and case (C_1) occurs, then (Figure 4a) by the switching of R in Γ' it has $\dot{g}'_{cd} = \dot{g}_{cd} + 1$, $\dot{g}'_{ck} = \dot{g}_{ck} + 1$ and $\dot{g}'_{dk} = \dot{g}_{dk}$;*
- *if R is a ρ_2-pair of color c in Γ and case (C_2) occurs, then, with the above notations, by the switching of R in Γ' it has $\dot{g}'_{ci} = \dot{g}_{ci}$, $\dot{g}'_{cj} = \dot{g}_{cj} + 1$ and $\dot{g}'_{dk} = \dot{g}_{dk}$ (Figures 4b and 4c).*

If Γ is a 3-colored graph and R is a ρ-pair either in Γ or in $\bar{\Gamma}$, then we will denote by Γ' the graph obtained by the switching of R, by χ the Euler characteristic of Γ and by χ' the Euler characteristic of Γ'.

Theorem 3.3 *Let Γ be a 3-colored graph and let R be a ρ_1-pair either in Γ or in $\bar{\Gamma}$. Then χ' is equal either to χ or $\chi + 1$.*

If R is a ρ_2-pair either in Γ or in $\bar{\Gamma}$ and case (C_1) occurs, then χ' is equal either to $\chi + 2$ or $\chi + 3$.

If R is a ρ_2-pair either in Γ or in $\bar{\Gamma}$ and case (C_2) occurs, then χ' is equal either to $\chi + 1$ or $\chi + 2$.

Proof. If $R = (\mathbf{e}, \mathbf{f})$ is a ρ_1-pair in Γ, then $\Gamma_{cd}(\mathbf{e}) \neq \Gamma_{cd}(\mathbf{f})$.

If at least one of $\Gamma_{cd}(\mathbf{e})$ or $\Gamma_{cd}(\mathbf{f})$ is a cycle, then, by the above Remark, $\chi' = \chi$.

If both $\Gamma_{cd}(\mathbf{e})$ and $\Gamma_{cd}(\mathbf{f})$ are paths or R is a ρ_1-pair in $\bar{\Gamma}$(hence at least one of $\Gamma_{cd}(\mathbf{e})$ or $\Gamma_{cd}(\mathbf{f})$ is a cycle), then $\chi' = \chi + 1$.

If $R = (\mathbf{e}, \mathbf{f})$ is a ρ_2-pair in Γ and case (C_1) occurs, then, by the above Remark, $\chi' = \chi + 2$.

If $R = (\mathbf{e}, \mathbf{f})$ is a ρ_2-pair in $\bar{\Gamma}$ and case (C_1) occurs, then, by the above Remark, $\chi' = \chi + 3$ (as a consequence of $(\mathbf{u}, \mathbf{v})$ being a wound).

If $R = (\mathbf{e}, \mathbf{f})$ is a ρ_2-pair in Γ and case (C_2) occurs, then, by the above Remark, $\chi' = \chi + 1$.

If $R = (\mathbf{e}, \mathbf{f})$ is a ρ_2-pair in $\bar{\Gamma}$ and case (C_2) occurs, then, by the above Remark, $\chi' = \chi + 2$ (as a consequence of $(\mathbf{u}, \mathbf{v})$ being a wound). □

4. Closed surfaces

The present section is devoted to present the topological effects of the switching of a ρ-pair in a 3-colored graph Γ, without boundary.

Recall that Γ always represents a closed surface $|K(\Gamma)|$ (which is orientable iff Γ is bipartite).

Moreover, for each color $i \in \Delta_2$, the subgraph $\Gamma_{\hat{\imath}}$ has bicolored cycles as components. Hence, even in the case of Γ being non-bipartite, the $\hat{\imath}$-residues of Γ are bipartite.

Furthermore, if $|K(\Gamma)|$ is the closed surface represented by Γ, then $\chi(|K(\Gamma)|) = \chi(\Gamma)$.

As above, we will denote simply by χ the Euler characteristic of both Γ and $|K(\Gamma)|$.

We recall that, if $R = (\mathbf{e}, \mathbf{f})$ is any ρ_1-pair in Γ, then, by Definition 3.7, there exists a unique way to switch R, i.e. by preserving the bipartition of $V(\Gamma)$ induced by $\Gamma_{\hat{d}}$.

Let now Γ be a connected, non-bipartite, 3-colored graph. Let $R = (\mathbf{e}, \mathbf{f})$ be a ρ_2-pair in Γ, involving color c.

Label arbitrarily by $\mathbf{e_0}$ and $\mathbf{e_1}$ the endpoints of $\mathbf{e}$ and consider the bipartitions of $V(\Gamma)$ induced by $\Gamma_{\hat{d}}$ (resp. $\Gamma_{\hat{k}}$), that is, setting V_j^d (resp. V_j^k), $j = 0, 1$, so that $\mathbf{e_j} \in V_j^d$ (resp. $\mathbf{e_j} \in V_j^k$), the set of vertices of the above bipartition, then $V(\Gamma) = V_0^d \cup V_1^d$ (resp. $V(\Gamma) = V_0^k \cup V_1^k$).

As above, suppose that the graphs $\Gamma_{\hat{d}}$ and $\Gamma_{\hat{k}}$ are directed so that $\mathbf{e_0}$ (resp. $\mathbf{e_1}$) is the tail (resp. the head) of $\mathbf{e}$.

Consider the cycles $\Gamma_{\hat{d}}(\mathbf{e})$ and $\Gamma_{\hat{k}}(\mathbf{e})$ with the orientations induced by $\mathbf{e}$. For $i = d, k$, since R is a ρ_2-pair, $\Gamma_{\hat{\imath}}(\mathbf{e}) = \Gamma_{\hat{\imath}}(\mathbf{f})$, and the orientation of $\mathbf{e}$ induces an orientation on the edge $\mathbf{f}$: let, as above, $\mathbf{f_0^i}$ be the first end of $\mathbf{f}$ which we meet starting from $\mathbf{e_1}$ along $\Gamma_{\hat{\imath}}(\mathbf{e})$. Since Γ is non-bipartite, both cases (C_1) and (C_2) can occur, whereas, because of Remark 3.2, if Γ is bipartite, then only case (C_1) of Section 3 can occur.

From now on, apart from denoting as above by χ the Euler characteristic of Γ (and of $|K(\Gamma)|$), by p the order of Γ and by g_B the number of B - residues of Γ, we will denote by χ' the Euler characteristic of the graph Γ' (and of $|K(\Gamma')|$), obtained by the switching of a ρ-pair in Γ, , by p' the order of Γ' and by g'_B the number of B - residues of Γ'. Finally, for each graph Ψ we will denote by $gen(|K(\Psi)|)$ both the genus of Ψ and of $|K(\Psi)|$.

Theorem 4.1 *Let Γ be a connected, 3-colored graph representing the closed surface $|K(\Gamma)|$.*

Let $R = (\mathbf{e}, \mathbf{f})$ be a ρ-pair in Γ and let Γ' be the graph obtained from Γ, by the switching of R.

Then:

(a) if R is a ρ_1-pair, then $|K(\Gamma)| \simeq |K(\Gamma')|$;

(b^I) if R is a ρ_2-pair and case (C_1) occurs, then:

- *(b^I_1) if Γ' is connected, $|K(\Gamma)| \simeq K(\Gamma')|\#\mathbb{H}$;*
- *(b^I_2) if Γ' has two connected components, say Γ_1 and Γ_2, then $|K(\Gamma)| \simeq |K(\Gamma_1)|\#|K(\Gamma_2)|$.*

(b^{II}) if R is a ρ_2-pair, and case (C_2) occurs, hence Γ is non-bipartite, then: Γ' is connected and $|K(\Gamma)| \simeq |K(\Gamma')|\#N_1$;

Proof. (a) As pointed out in Theorem 3.2, Γ' and Γ are both connected; moreover, since both $\Gamma_{cd}(\mathbf{e})$ and $\Gamma_{cd}(\mathbf{f})$ are cycles, then $\chi' = \chi$ (see Theorem 3.3), hence $|K(\Gamma')| \simeq |K(\Gamma)|$.

(b^I) By switching a ρ_2-pair (case (C_1)), we obtain the new graph Γ' with $\chi' = \chi + 2$, by Theorem 3.3.

Now, we distinguish the case of R being an edge-cut (case b^I_2) or not (case b^I_1).

(b^I_1) If Γ' is connected, then $gen(K(|\Gamma'|)) = gen(K(|\Gamma|)) - 1$ (or $gen(K(|\Gamma'|)) = gen(K(|\Gamma|)) - 2$), if $K(|\Gamma|$ is orientable (resp. non-orientable), hence $|K(\Gamma)| \simeq |K(\Gamma')|\#\mathbb{H}$.

(b^I_2) If Γ' has two connected components Γ_1 and Γ_2 (the only possible case, by theorem 2), then $|K(\Gamma)|$ splits into two surfaces $|K(\Gamma_1)|$ and $|K(\Gamma_2)|$, and $gen(|K(\Gamma_1)|) + gen(|K(\Gamma_2)|) = gen(|K(\Gamma|)$.

(b^{II}) Since, in this case, $|K(\Gamma)|$ is non-orientable, hence Γ is non-bipartite and $\chi' = \chi + 1$.

If Γ' is bipartite and connected, then $gen(|K(\Gamma')|) = \frac{1}{2}(gen(|K(\Gamma)|) - 1)$, hence $gen(|K(\Gamma)|)$ is necessarily odd and $|K(\Gamma)| \simeq |K(\Gamma')|\#N_1$.

If Γ' is non-bipartite and connected, then $gen(|K(\Gamma')|) = (gen(|K(\Gamma)|) - 1)$, hence $|K(\Gamma)| \simeq |K(\Gamma')|\#N_1$. $\square$

Corollary 4.1 *Let (Σ, σ) be a 3-colored graph representing the 2-sphere $\mathbb{S}^2$, $R = (\mathbf{e}, \mathbf{f})$ be a ρ_2-pair of Σ and let (Σ', σ') be the graph obtained by switching R. Then Σ' has two connected components, both representing $\mathbb{S}^2$.*

Proof. Since (Σ, σ) represents $\mathbb{S}^2$, $\chi(\Sigma) = 2$, hence $\chi(\Sigma') = 4$. This implies that Σ' has two connected components, both having Euler characteristic 2, and so representing $\mathbb{S}^2$. □

Remark 4.1 *Note that, if (Γ, γ) is a 3-colored graph we have:*

(1) if Γ is a crystallization, then all pairs of edges colored c lie on the same (unique) $\{c, j\}$-colored cycle, for each color $j \in \{d, k\}$, so all pairs of c-colored edges are ρ_2-pairs;

(2) if Γ is not contracted, then there is at least a pair $(\mathbf{e}, \mathbf{f})$ of edges colored c, say, lying on the same $\{c, k\}$-colored cycle, but on different $\{c, d\}$-colored cycles; hence, $R = (\mathbf{e}, \mathbf{f})$ is a ρ_1-pair.

Remark 4.2 *We recall that, if M is an n-manifold and the $(n+1)$-colored graph Γ represents M, then Γ is called rigid if it has neither $\rho_{(n-1)}$-pairs, nor ρ_n-pairs (see [5]).*

Because of Remark 4.1, there exist no 3-colored graphs with this property, hence rigid.

5. Surfaces with boundary

The present section is devoted to describe the effects of switching ρ- and ρ^*-pairs in 3-colored connected graphs with non-empty boundary. Hence, in the following, Γ will be a 3-colored, connected graph with boundary.

As pointed out before, given a 3-colored graph with boundary (Γ, γ) , for each color $i \in \Delta_2$, the subgraph $\Gamma_{\hat{\imath}}$ has bicolored cycles or paths as components. Hence, the $\hat{\imath}$-residues of Γ are always bipartite.

Given the graph Γ, let R be a ρ-pair in Γ (resp. in $\bar{\Gamma}$) and Γ' be the graph obtained by the switching of R.

Moreover, when Γ' splits into two connected components, we denote them by Γ'_1 and by Γ'_2 respectively; they are graphs representing surfaces of genus g_1, g_2 (resp. h_1, h_2) having λ_1 and λ_2 holes respectively.

Note that, when $k = 2$, if $\Gamma_{c2}(\mathbf{e}) = \Gamma_{c2}(\mathbf{f})$ contains boundary vertices, then the switching does not alter the boundary.

Conversely, the switching can alter the boundary only if $c = 2$ or $d = 2$ and boundary vertices are ends of paths passing through $\mathbf{e}$ or $\mathbf{f}$.

In the following, we set $\Xi = \Gamma_{cd}(\mathbf{e}) \cup \Gamma_{cd}(\mathbf{f})$, if R is a ρ_1-pair and $\Xi = \Gamma_{cd}(\mathbf{e}) \cup \Gamma_{ck}(\mathbf{e}) = \Gamma_{cd}(\mathbf{f}) \cup \Gamma_{ck}(\mathbf{f})$, if R is a ρ_2-pair.

Theorem 5.1 *Let Γ be a 3-colored graph with boundary, and let Γ' be the graph obtained by the switching of a ρ-pair R.*

ρ_1- pairs

1.1 If R is a ρ_1-pair in Γ and $\Xi = \Gamma_{cd}(\mathbf{e}) \cup \Gamma_{cd}(\mathbf{f})$ has either 0 or 2 boundary vertices, then $\chi = \chi'$ and $\lambda' = \lambda$.

1.2 If R is a ρ_1-pair in Γ and $\Xi = \Gamma_{cd}(\mathbf{e}) \cup \Gamma_{cd}(\mathbf{f})$ has 4 boundary vertices, then $\chi' = \chi + 1$ and, either $\lambda' = \lambda + 1$, or $\lambda' = \lambda$, or $\lambda' = \lambda - 1$.

ρ_2- pairs

2.1 If R is a ρ_2-pair in Γ and case (C_1) occurs, then $\chi' = \chi + 2$ and $\lambda' = \lambda$.

2.2 If R is a ρ_2-pair in Γ and case (C_2) occurs, then $\chi' = \chi + 1$ and $\lambda' = \lambda$.

Proof. In all cases, for Euler characteristic calculation, we refer to Theorem 3.3.

1.1 If R is a ρ_1-pair in Γ and $\Xi = \Gamma_{cd}(\mathbf{e}) \cup \Gamma_{cd}(\mathbf{f})$ contains no boundary vertices, the switching of R has no influence on the boundary graph and, therefore, on the number of connected components of Γ', hence $\lambda' = \lambda$.

If the boundary vertices of Ξ are exactly two, then they are the ends, in Γ, either of a $(c, 2)$-colored path C or of a $(d, 2)$-colored path C' and correspond in $\partial\Gamma$ either to a pair of c-adjacent vertices or to a pair of d-adjacent vertices. Therefore, by switching R, either C (resp. C') remains unaltered, or it splits into two $(c, 2)$ (resp. $(d, 2)$)-colored connected components: one of them is a cycle and the other one is a path with the same ends as before.

1.2 Suppose that Ξ has four boundary vertices and suppose, w.l.o.g., that $\mathbf{x}, \mathbf{y}$ (resp. $\mathbf{z}, \mathbf{w}$) are the boundary vertices of $\Gamma_{cd}(\mathbf{e})$ (resp. $\Gamma_{cd}(\mathbf{f})$) such that $\mathbf{e}$ (resp. $\mathbf{f}$) lies on the $(i, 2)$-colored path with ends $\mathbf{x}, \mathbf{y}$ (resp. $\mathbf{z}, \mathbf{w}$), where either $i = d$ or $i = c$. Suppose, moreover, that, once the orientation on the paths induced by that of $\mathbf{e}$ and $\mathbf{f}$ (the last one, relatively to $\Gamma_{\hat{d}}$) is fixed, $\mathbf{x}$ is the vertex preceding $\mathbf{e_0}$, while $\mathbf{y}$ is the vertex following $\mathbf{e_1}$ along $\Gamma_{cd}(\mathbf{e})$, and $\mathbf{z}$ is the vertex preceding $\mathbf{f_0^d}$ while $\mathbf{w}$ is the vertex following $\mathbf{f_1^d}$ along $\Gamma_{cd}(\mathbf{f})$.

Fix a color i in $\{c, d\}, i \neq 2$. Consider the i-colored edges in $\partial\Gamma$, say $\mathbf{a}$ and $\mathbf{b}$, joining $\mathbf{x}$ with $\mathbf{y}$ and $\mathbf{z}$ with $\mathbf{w}$. The switching of R gives rise, in $\partial\Gamma'$, to two i-colored edges, say $\mathbf{a'}$ and $\mathbf{b'}$, joining $\mathbf{x}$ with $\mathbf{w}$ and $\mathbf{y}$ with $\mathbf{z}$.

In fact, both in the case of Γ' being connected and in the case of Γ' splitting into two connected components, if $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}$ lie on the same boundary component ${}^{\partial}C$ of $\partial\Gamma$ (by using the above notations, ${}^{\partial}C$ is a (k, i)-colored cycle) and if, by going through ${}^{\partial}C$ according to one of the possible orientations of ${}^{\partial}C$, we meet $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}$, in this order, then ${}^{\partial}C$ splits into two connected components, and $\lambda' = \lambda + 1$.

If, conversely, we meet $\mathbf{x}, \mathbf{y}, \mathbf{w}, \mathbf{z}$, in this order, then Γ' is connected, ${}^{\partial}C$ is unaltered by the switching, and $\lambda' = \lambda$.

Finally, if $\mathbf{x}, \mathbf{y}$ lie on a boundary component ${}^{\partial}C'$ of $\partial\Gamma$ and $\mathbf{z}, \mathbf{w}$ lie on a different boundary component ${}^{\partial}C''$ of $\partial\Gamma$, then Γ' is connected and, since after the switching, in $\partial\Gamma'$, the vertex $\mathbf{x}$ is adjacent to the vertex $\mathbf{w}$ and the vertex $\mathbf{y}$ is adjacent to the vertex $\mathbf{z}$, then the vertices $\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}$ belong to the same connected component of $\partial\Gamma'$, and $\lambda' = \lambda - 1$.

2.1 and 2.2 Note that in this case the switching don't alter the boundary. □

Theorem 5.2 *(a) if R is a ρ_1-pair of Γ or $\bar{\Gamma}$, and Γ' is connected, then we have one among the following cases:*

$|K(\Gamma)| \simeq |K(\Gamma')|$ *or* $|K(\Gamma)| \simeq |K(\Gamma')| \# N_1$ *or* $|K(\Gamma)| \simeq |K(\Gamma')| \# \mathbb{H} \#_\partial T_{0,1}$ *or* $|K(\Gamma)| \simeq |K(\Gamma')| \#_\partial T_{0,1}$.

(b) if R is a ρ_1-pair of Γ, and Γ' splits into two connected components Γ_1 and Γ_2, then $|K(\Gamma_1)| \# |K(\Gamma_2)| \simeq |K(\Gamma)| \#_\partial T_{0,1}$

Proof. (a) Via Theorems 3.3 and 5.1, if R is a ρ_1-pair in Γ and Γ' is connected, we can have the following cases:

- $\chi = \chi'$ and $\lambda' = \lambda$, then $|K(\Gamma')| \simeq |K(\Gamma)|$;
- $\chi' = \chi + 1$ and $\lambda' = \lambda$, hence $|K(\Gamma')|$ is non-orientable and its genus is $gen|K(\Gamma)| - 1$. Then $|K(\Gamma)| \simeq |K(\Gamma')| \# N_1$;
- $\chi' = \chi + 1$ and $\lambda' = \lambda + 1$, hence $gen|K(\Gamma')| = gen|K(\Gamma)| - 1$ (resp. $gen|K(\Gamma)| - 2$) if it is orientable (resp. non-orientable) and it has one more hole. Then $|K(\Gamma)| \simeq |K(\Gamma')| \# \mathbb{H} \#_\partial T_{0,1}$;

If R is a ρ_1-pair in $\bar{\Gamma}$ and Γ' is connected, then R is a wound, only the first case can occur and $|K(\Gamma')| \simeq |K(\Gamma)| \#_\partial T_{0,1}$.

(b) Let Γ' have two connected components, say Γ_1 and Γ_2 with λ_1 and λ_2 boundary components, respectively.

Then R is an edge-cut with respect to color d, and case $\lambda' = \lambda - 1$ cannot occur.

If $\lambda' = \lambda + 1$, it is easy to see that $gen(|K(\Gamma_2)|) = \frac{1}{2}(gen(|K(\Gamma')|) - gen(|K(\Gamma_1)|) + 1)$ (resp. $gen(|K(\Gamma_2)|) = gen(|K(\Gamma')|) - gen(|K(\Gamma_1)|) + 1)$ if $|K(\Gamma_2)|$ is orientable (resp. if it is non-orientable); whereas, if $\lambda' = \lambda$, then at least one of $|K(\Gamma_1)|$, $|K(\Gamma_2)|$ is non-orientable (w.l.o.g. we set $|K(\Gamma_2)|$ non-orientable) and

$gen(|K(\Gamma_2)|) = gen(|K(\Gamma')|) - gen(|K(\Gamma_1)|)$, if $|K(\Gamma_1)|$ is non-orientable too,

$gen(|K(\Gamma_2)|) = gen(|K(\Gamma')|) - 2\, gen(|K(\Gamma_1)|)$ otherwise.

Note that, by definition, after a switching of a ρ_1-pair in $\bar{\Gamma}$, the obtained graph Γ' is necessarily connected, then $|K(\Gamma')| \simeq |K(\bar{\Gamma})| \simeq |K(\Gamma)| \#_\partial T0, 1$. This proves the statement. □

Now, we analyze the effects of the switching of a ρ_2-pair.

Note that, if R is a ρ_2-pair and case (C_2) occurs, then Γ', obtained by the switching of R, is connected.

Theorem 5.3 *(a) If R is a ρ_2-pair of color c of Γ and case (C_1) holds, and Γ' is connected, then: $|K(\Gamma)| \simeq |K(\Gamma')|\#\mathbb{H}$.*

(b) If R is a ρ_2-pair of color c of Γ or $\bar{\Gamma}$ and case (C_2) holds, then Γ' is connected, and $|K(\Gamma)| \simeq |K(\Gamma')|\#N_1$.

(c) If R is a ρ_2-pair in $\bar{\Gamma}$ (hence $c = 2$) and case (C_1) holds, and Γ' is connected, then : $|K(\Gamma)| \simeq T_{g,\lambda} \simeq |K(\Gamma')|\#\mathbb{H}\#_\partial T_{0,1}$.

(d) if R is a ρ_2-pair of color c in Γ and case (C_1) holds, and Γ' has two connected components, say Γ_1 and Γ_2, then $|K(\Gamma) \simeq |K(\Gamma_1)|\#|K(\Gamma_2)|$.

(e) if R is a ρ_2-pair of color c in $\bar{\Gamma}$, case (C_1) holds, and Γ' has two connected components, say Γ_1 and Γ_2, then $|K(\Gamma)| \simeq |K(\Gamma_1)|\#|K(\Gamma_2)|\#_\partial T_{0,1}$.

Proof. First, let us note that, if R is a ρ_2-pair of color c in Γ and case (C_1) occurs, then $\chi' = \chi + 2$, if R is a ρ_2-pair of color c, case (C_2), then $\chi' = \chi + 1$.

If R is a ρ_2-pair of $\bar{\Gamma}$, then Ξ ($= \Gamma_{cd}(\mathbf{e}) \cup \Gamma_{ck}(\mathbf{e})$) has exactly two boundary vertices (exactly $\mathbf{u}$ and $\mathbf{v}$) in Γ and no boundary vertices in $\bar{\Gamma}$. Actually, in this case, $\mathbf{u}$ and $\mathbf{v}$ are the only boundary vertices of a connected component of $\partial\Gamma$ and they become internal vertices in $\bar{\Gamma}$ (hence in Γ') and $\lambda' = \lambda - 1$.

Hence, in case (C_1), $\chi' = \chi + 3$ (see Theorem 3.3) and $\lambda' = \lambda - 1$.

In case (C_2), $\chi' = \chi + 2$ and $\lambda' = \lambda - 1$.

Moreover, if Ξ ($= \Gamma_{cd}(\mathbf{e}) \cup \Gamma_{ck}(\mathbf{e})$) contains either 0 or 2 boundary vertices, then the switching of R cannot change the boundary, hence $\partial\Gamma$ and $\partial\Gamma'$ are color - isomorphic; in particular, they have the same number of connected components.

On the other hand, if Ξ has 4 boundary vertices, necessarily $c = 2$ (hence this holds if R is a ρ_2-pair both of Γ and of $\bar{\Gamma}$) and Γ' is connected, then there exist in Γ a $(d, 2)$-colored path C and a $(k, 2)$-colored path C', containing $\mathbf{e}$ and $\mathbf{f}$ and the switching of R alters the paths, but preserves the adjacencies in the boundary graph, hence $\lambda' = \lambda$. □

References

[1] P. Bandieri, P. Cristofori, C. Gagliardi: *Nonorientable 3-manifolds admitting coloured triangulations with at most 30 tetrahedra.* J. Knot Theory Ramifications **18** (2009) 381–395.

[2] P. Bandieri, P. Cristofori, C. Gagliardi: *A census of genus two 3-manifolds up to 42 coloured tetrahedra.* Discrete Math. **310** (2010) 2469–2481.

[3] P. Bandieri, M. R. Casali, C. Gagliardi: *Representing manifolds by crystallization theory: foundations, improvements and related results.* Atti Sem. Mat. Fis. Univ. Modena. suppl. **49** (2001) 283–337.

[4] P. Bandieri, M. R. Casali, P. Cristofori, L. Grasselli, M. Mulazzani: *Computational aspects of crystallization theory: complexity, catalogues and classifications of 3-manifolds.* Atti Sem. Mat. Fis. Univ. Modena **58** (2011) 11–45.

[5] P. Bandieri, C. Gagliardi: *Rigid gems in dimension n.* Boletin de la Sociedad Matematica Mexicana **18** (2012) 55–67.

[6] B. A. Burton: *Enumeration of non-orientable 3-manifolds using face-pairing graphs and union-find. Discrete Comput. Geom.* **38 (3)** (2007) 527–571.

[7] M. R. Casali: *Catalogues of PL-manifolds and complexity estimations via crystallization theory.* Oberwolfach Report no. **24** (2012) 58–61. [DOI: 10.4171/OWR/2012/24]

[8] M. R. Casali, P. Cristofori: *A catalogue of orientable 3-manifolds triangulated by* 30 *coloured tetrahedra.* J. Knot Theory Ramifications **17** (2008) 1–23.

[9] M. R. Casali, P. Cristofori: *Coloured graphs representing PL 4-manifolds.* Electronic Notes in Discrete Mathematics **40** (2013) 83–87.

[10] M. R. Casali, P. Cristofori: *Cataloguing PL 4-manifolds by gem-complexity.* arXiv:1408.0378.

[11] J. Fabrega, M. A. Fiol: *Connectivity: properties and structure.* In Discrete Mathematics and its Applications, Handbook of Graph Theory, Eds. J. L. Gross and J. Yellen, CRC Press (2004) 194–213.

[12] M. Ferri, C. Gagliardi, L. Grasselli: *A graph-theoretical representation of PL-manifolds. A survey on crystallizations.* Aeq. Math. **31** (1986) 121–141.

[13] M. Ferri, S. Lins: *Topological aspects of edge-fusions in* 4*-graphs.* J. Comb. Theor. Ser. B **51** (1991) 227–243.

[14] C. Gagliardi: *On a class of* 3*-dimensional polyhedra.* Ann. Univ. Ferrara **33** (1987) 51–88.

[15] C. Gagliardi, G. Volzone: *Handles in graphs and sphere bundles over* $\mathbb{S}^1$. Europ. J. Combinatorics **8** (1987) 151–158.

[16] L. C. Glaser: *Geometrical combinatorial topology.* Van Nostrand Reinhold Math. Studies, New York, 1987.

[17] J. L. Gross, T.W. Tucker: *Topological graph theory.* John Wiley & Sons, New York, 1987.

[18] P. Heegaard: *Forstudier til topologisk teori för de algebraiske Sammenhäeng.* Nordiske Forlag Ernst Bojesen, Copenhagen (1898); french translation: *Bull. Soc. Math. France* **44** (1916) 161–212.

[19] P. J. Hilton, S. Wylie: *An introduction to algebraic topology-homology theory.* Cambridge Univ. Press, Cambridge, 1960.

[20] S. Lins: *Gems, computers and attractors for* 3*-manifolds.* Knots and Everything **5**, 1995.

[21] S. Matveev: *Complexity theory of three-dimensional manifolds. Acta Appl. Math.* **19** (1990) 101–130.

[22] M. Pezzana: *Sulla struttura topologica delle varietà compatte.* Atti Sem. Mat. Fis. Univ. Modena, **23** (1974) 269–277.

[23] M. Pezzana: *Diagrammi di Heegaard e triangolazione contratta*. Boll. Un. Mat. Ital. **12 (4)** (1975) 98–105.

[24] C. Rourke, B. Sanderson: *Introduction to piecewise-linear topology*. Springer Verlag, New York-Heidelberg, 1972.

[25] A. T. White: *Graphs, groups and surfaces*. North Holland, 1984.

Fixed points, bounded orbits and attractors of planar flows

Héctor Barge and José M. R. Sanjurjo*

Departamento de Geometría y Topología
Universidad Complutense de Madrid
28040 Madrid, Spain
hbarge@ucm.es   jose_sanjurjo@mat.ucm.es

Affectionately dedicated to José María Montesinos-Amilibia.

ABSTRACT

In this paper we provide a dynamical characterization of isolated invariant continua which are global attractors for planar dissipative flows. As a consequence, a sufficient condition for an isolated invariant continuum to be either an attractor or a repeller is derived for general planar flows.

2010 Mathematics Subject Classification: 34C35, 34D23, 37C10, 37C25, 37C70.
Key words: Attractor, Fixed point, Bounded orbit, Dissipative flow, Isolated invariant set.

1. Introduction

In this paper we are concerned with the study of planar flows $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$. In particular, we provide a dynamical characterization of isolated invariant continua which are global attractors for planar dissipative flows. This characterization is inspired by a result of Alarcón, Guíñez and Gutiérrez about dissipative planar embeddings with only one fixed point (see [1]). Moreover we will derive a sufficient condition for a planar continuum to be an attractor or a repeller provided that it contains all the fixed points of φ.

We shall use through the paper the standard notation and terminology in the theory of dynamical systems. In particular we shall use the notation $\gamma(x)$ for the *trajectory* of the point x, i.e. $\gamma(x) = \{xt \mid t \in \mathbb{R}\}$. By the *omega-limit* of a point x

*The authors are partially supported by project MTM2012-30719.

we understand the set $\omega(x) = \bigcap_{t>0} \overline{x[t,\infty)}$ while the *negative omega-limit* is the set $\omega^*(x) = \bigcap_{t<0} \overline{x(-\infty,t]}$. An invariant compactum K is *stable* if every neighborhood U of K contains a neighborhood V of K such that $V[0,\infty) \subset U$. Similarly, K is *negatively stable* if every neighborhood U of K contains a neighborhood V of K such that $V(-\infty,0] \subset U$. An invariant compactum K is said to be *attracting* provided that there exists a neighborhood U of K such that $\omega(x) \subset K$ for every $x \in U$. In an analogous way, K is said to be *repelling* provided that there exists a neighborhood U of K such that $\omega^*(x) \subset K$ for every $x \in U$. An *attractor* (or *asymptotically stable* compactum) is an attracting stable set and a *repeller* is a repelling negatively stable set.

If K is an attracting set, its region (or basin) of attraction $\mathcal{A}$ is the set of all points $x \in M$ such that $\omega(x) \subset K$. An attracting set K is *globally attracting* provided that $\mathcal{A}$ is the whole phase space. If K is an attractor and $\mathcal{A}$ is the whole phase space, then K is said to be a *global attractor* (or *globally asymptotically stable* compactum). For the reader interested in a detailed treatment of attracting sets we recommend [14] and [18].

Through this paper we shall deal with a special kind of invariant compacta, the so-called *isolated invariant sets* (see [5, 6, 7, 17] for details). These are compact invariant sets K which possess an *isolating neighborhood*, i.e. a compact neighborhood N such that K is the maximal invariant set in N. For instance, attractors and repellers are isolated invariant sets. We shall make use of the next result which states that isolated globally attracting continua for planar flows are stable.

Theorem 1.1 (Morón, Sánchez-Gabites and Sanjurjo [14]) *Every connected isolated globally attracting set K in $\mathbb{R}^2$ is a global attractor.*

A special kind of isolating neighborhoods shall be useful in the sequel, the so-called *isolating blocks*, which have good topological properties. More precisely, an isolating block N is an isolating neighborhood such that there are compact sets $N^i, N^o \subset \partial N$, called the entrance and the exit sets, satisfying

1. $\partial N = N^i \cup N^o$;
2. for each $x \in N^i$ there exists $\varepsilon > 0$ such that $x[-\varepsilon,0) \subset M - N$ and for each $x \in N^o$ there exists $\delta > 0$ such that $x(0,\delta] \subset M - N$;
3. for each $x \in \partial N - N^i$ there exists $\varepsilon > 0$ such that $x[-\varepsilon,0) \subset \mathring{N}$ and for every $x \in \partial N - N^o$ there exists $\delta > 0$ such that $x(0,\delta] \subset \mathring{N}$.

These blocks form a neighborhood basis of K in M. If the flow is differentiable, the isolating blocks can be chosen to be differentiable manifolds which contain N^i and N^o as submanifolds of their boundaries and such that $\partial N^i = \partial N^o = N^i \cap N^o$. In particular, for flows defined on $\mathbb{R}^2$, the exit set N^o is a disjoint union of a finite number of intervals $J_1, \ldots, J_m$ and circumpherences $C_1, \ldots, C_n$ and the same is true for the entrance set N^i.

The dynamical structure near isolating invariant sets shall play an important role in this paper and it is described by the

Theorem 1.2 (Ura-Kimura-Egawa [20, 8]) *Let M be a locally compact separable metric space and φ a flow on M. Suppose $K \neq M$ is a non-empty isolated invariant compactum. Then, one and only one of the following alternatives holds:*

1. *K is an attractor;*

2. *K is a repeller;*

3. *There exist points $x \in M - K$ and $y \in M - K$ such that $\emptyset \neq \omega(x) \subset K$ and $\emptyset \neq \omega^*(y) \subset K$.*

We shall also make use of a classical result of C. Gutiérrez about smoothing of 2-dimensional flows.

Theorem 1.3 (Gutiérrez [10]) *Let $\varphi : M \times \mathbb{R} \to M$ be a continuous flow on a compact C^∞ two-manifold M. Then there exists a C^1 flow ψ on M which is topologically equivalent to φ. Furthermore, the following conditions are equivalent:*

1. *any minimal set of φ is trivial;*

2. *φ is topologically equivalent to a C^2 flow;*

3. *φ is topologically equivalent to a C^∞ flow.*

By a trivial minimal set we understand a fixed point, a closed trajectory or the whole manifold if M is the 2-dimensional torus and φ is topologically equivalent to an irrational flow. We readily deduce from Gutiérrez' Theorem applied to the Alexandrov compactification of the plane that continuous flows $\varphi : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ are topologically equivalent to C^∞ flows.

Some basic results about planar vector fields such as, the Poincaré-Bendixson Theorem, the Tubular Flow Theorem and the elementary properties of transversal sections shall be used through the paper. Two good references covering this material are the book of Hirsch, Smale and Devaney [12] and the monograph of Palis and de Melo [15]. In addition, a form of homotopy theory, namely *shape theory*, which is the most suitable for the study of global topological properties in dynamics, will be occasionally used. Although shape theory is not necessary to understand the paper, we recommend to the reader the references [4], which contains an exhaustive treatment of the subject and [19], which covers some dynamical applications of this theory.

2. Planar dissipative systems and isolated invariant continua

We start this section by recalling the definition of *dissipative flow*. Let M be a locally compact metric space and $\varphi : M \times \mathbb{R} \to M$ a flow on M. The flow φ is said to be *dissipative* if $\omega(x) \neq \emptyset$ for every $x \in M$ and $\bigcup_{x\in M} \omega(x)$ has compact closure. If the phase space M is not compact, dissipativeness is equivalent to $\{\infty\}$ being a repeller of the extended flow $\widehat{\varphi} : (M \cup \{\infty\}) \times \mathbb{R} \to M \cup \{\infty\}$ to the Alexandrov compactification of M leaving ∞ fixed (See[9, 11, 19]), and therefore to the existence of a global attractor for φ.

The following result gives a relation between global asymptotic stability of a fixed point and the non-existence of additional fixed points in the case of discrete dynamical systems.

Theorem 2.1 (Alarcón-Guíñez-Gutiérrez [1], Ortega-Ruiz del Portal [16]) *Assume that $h \in \mathcal{H}_+$ (orientation preserving homeomorphisms of $\mathbb{R}^2$) is dissipative and p is an asymptotically stable fixed point of h. The following conditions are equivalent:*

1. *p is globally asymptotically stable;*
2. *$Fix(h) = p$ and there exists an arc $\gamma \subset S^2$ with end points at p and ∞ such that $h(\gamma) = \gamma$.*

The proof in [1] is based on Brouwer's theory of fixed point free homeomorphisms of the plane. Ortega and Ruiz del Portal give in [16] an alternative proof based on the theory of prime ends.

Inspired by Theorem 2.1, the authors prove in [2] that for continuous and dissipative dynamical systems the result is satisfied even if the fixed point p is substituted by a connected attractor K which contains every fixed point of the flow. We prove in our next result that the asymptotical stability condition can be dropped from the hypothesis, obtaining in this way a simple characterization of global attractors of dissipative planar flows.

Theorem 2.2 *Let K be an isolated invariant continuum of a dissipative flow φ in $\mathbb{R}^2$. The following conditions are equivalent:*

1. *K is a global attractor;*
2. *There are no fixed points in $\mathbb{R}^2 - K$ and there exists an orbit γ connecting ∞ and K (i.e. such that $||\gamma(t)|| \to \infty$ when $t \to -\infty$ and $\omega(\gamma) \subset K$).*

Proof. By the Gutiérrez Theorem [10] we can assume that the flow φ is differentiable. Since φ is dissipative, given $x \in \mathbb{R}^2$ its ω-limit is non-empty and compact. Moreover, by the Poincaré-Bendixson Theorem either $\omega(x)$ contains fixed points and, hence, $\omega(x) \cap K \neq \emptyset$ or $\omega(x)$ is a periodic orbit. If $\omega(x)$ is a periodic orbit then K is not contained in its interior since, in that case, γ would meet $\omega(x)$, which is impossible. Therefore, if $\omega(x)$ is not contained in K, then K is in the exterior of $\omega(x)$ and, moreover, $\omega(x)$ being a periodic orbit, there must exist a fixed point in its interior. Hence this point belongs to K, which is a contradiction. We conclude that if $\omega(x)$ is a periodic orbit then $\omega(x) \subset K$.

If $\omega(x)$ is not a periodic orbit then $\omega(x) \cap K \neq \emptyset$ and we shall prove that, in fact, $\omega(x) \subset K$. We suppose, to get a contradiction, that there exists $y \in \omega(x) - K$. By hypothesis y is not a fixed point and, thus, we can take a local section I containing y and meeting transversally the trajectory of y. Since $y \notin K$ we can assume that $I \cap K = \emptyset$. It is a well-known fact that the trajectory of x meets I infinitely many times. We consider two consecutive points of intersection $x_1 = xt_1$ and $x_2 = xt_2$ with $x_1,x_2 \in I$, $0 < t_1 < t_2$ and $x[t_1, t_2] \cap I = \{x_1,x_2\}$. Then the set $C = x[t_1, t_2] \cup J$, where J is the subinterval of I bounded by x_1 and x_2, is a simple closed curve which, by the Jordan Theorem, decomposes $\mathbb{R}^2$ into two connected components U and V. If U is the bounded component then U is either positively or negatively invariant by [12]. Then, a simple argument involving again the Poincaré-Bendixson Theorem, leads to the existence of a fixed point in U which, by hypothesis, belongs to K. Now, the intersection of K with C is empty, which implies that $K \subset U \cup V$ and, K being connected, that $K \subset U$. If U is negatively invariant, the trajectory γ linking ∞ with K cannot enter in U since the only possibility would be through J, which is an exit set. This makes it impossible that $\omega(\gamma) \subset K$ and we get a contradiction with the hypothesis. If U is positively invariant then an easy argument shows that $y \in \omega(\gamma)$ in contradiction with the assumption. This proves that $\omega(x) \subset K$ for every $x \in \mathbb{R}^2$ and, as a consequence, K is a globally attracting set. Since K is isolated, by Theorem 1.1 K must be stable, i.e. a global attractor. This establishes the implication 2. $\Rightarrow$ 1. The converse implication is straightforward. $\square$

3. Attractors, repellers and bounded orbits

We present in this section a result which gives a readily testable sufficient condition for a planar compactum to be either an attractor or a repeller.

Theorem 3.1 *Let K be an isolated invariant continuum of a flow φ in $\mathbb{R}^2$. Suppose that there is a closed disk D containing K in its interior such that there are no fixed points in $D - K$ and that there is an orbit γ completely contained in $D - K$. Then K is either an attractor or a repeller. Moreover, K has trivial shape.*

Proof. We can assume again that φ is differentiable. Since $\overline{\gamma} \subset D$ we have that $\omega(\gamma) \subset D$ and $\omega^*(\gamma) \subset D$. We start by proving that there exists an orbit Γ in $D - K$ satisfying the additional condition that either $\omega(\Gamma) \subset K$ or $\omega^*(\Gamma) \subset K$. As a consequence of the Poincaré-Bendixson Theorem and the hypothesis of the present theorem we have that either $\omega(\gamma) \cap K \neq \emptyset$ or $\omega(\gamma)$ is a periodic orbit not meeting

K, and the same can be said for $\omega^*(\gamma)$. If $\omega(\gamma)$ is a periodic orbit not meeting K then K is in its interior and, by the Ura-Kimura Theorem, there exists a point x, also in the interior of $\omega(\gamma)$, with $\omega(x) \subset K$ or $\omega^*(x) \subset K$, and the same happens if $\omega^*(\gamma)$ is a periodic orbit not meeting K. Hence, in both cases Γ can be taken as the trajectory of x. On the other hand, we will prove that the possibility that both intersections, $\omega(\gamma) \cap K$ and $\omega^*(\gamma) \cap K$, are non-empty can never happen. Suppose, to get a contradiction, that $\omega(\gamma) \cap K \neq \emptyset$ and $\omega^*(\gamma) \cap K \neq \emptyset$. Take an isolating block N of K. By [6] N can be chosen to be a topological closed disk with i holes, one for every bounded component of $\mathbb{R}^2 - K$. We suppose that γ is in the unbounded component U (the argument being only slightly different in the other case) and consider the only circle $C \subset \partial N$ which is contained in U. Then, there exists a point $x \in C \cap \gamma$ leaving N and returning to N after a time $t \neq 0$, i.e. such that $xt \in C$ and $x(0,t) \cap N = \emptyset$. The possibility that the time t be positive or negative is irrelevant in this construction. Consider the arc A in C with extremes x and xt such that the topological circle $x[0,t] \cup A$ does not contain K in its interior. This arc can be mapped to the unit interval $I = [0,1]$ of the real line by a homeomorphism $h : A \to I$. If we take the point $x_1 \in \mathring{A}$ corresponding to the center of I then x_1 must leave N (in the past or in the future) and return again since, otherwise, the Theorem of Poincaré-Bendixson would imply the existence of a fixed point in the disk limited by $x[0,t] \cup A$. Hence, we can repeat the operation with $x_1[0,t_1] \cup A_1$, where A_1 is an arc in A with extremes x_1 and $x_1 t_1$ and the topological circle $x_1[0,t_1] \cup A_1$ does not contain K in its interior. Now take $x_2 \in \mathring{A}_1$ corresponding to the middle point of $h(A_1)$ and repeat the construction. In this way we obtain a sequence $A \supset A_1 \supset A_2 \supset \ldots$ of arcs whose intersection $\bigcap_{i=1}^{\infty} A_i$ consists of one point $p \in \partial N$. The orbit of p defines an internal tangency to ∂N, which is in contradiction with the properties of isolating blocks. We get from this contradiction that either $\omega(\gamma) \cap K = \emptyset$ or $\omega^*(\gamma) \cap K = \emptyset$ and, as a consequence, one of the two limits is a periodic orbit. Therefore, it follows from the remarks at the beginning of the proof that there exists an orbit Γ in $D - K$ satisfying the additional condition that either $\omega(\Gamma) \subset K$ or $\omega^*(\Gamma) \subset K$.

Suppose that $\omega(\Gamma) \subset K$. Then, $\omega^*(\Gamma)$ is a periodic orbit containing K in its interior. Let V be the interior of $\omega^*(\Gamma)$ and consider the flow restricted to $\overline{V}$. An elementary argument involving local sections again shows that $\omega^*(\Gamma)$ is a repeller for $\varphi|\overline{V}$ and, as a consequence, the restriction of φ to V is a dissipative flow. Then, using an arbitrary homeomorphism between V and $\mathbb{R}^2$ we can define a dissipative flow in $\mathbb{R}^2$ conjugated to $\varphi|V$ and satisfying the conditions of Theorem 2.2. We deduce from that theorem that K is an attractor of φ whose basin of attraction, V, is an open topological disk. Hence, K has trivial shape by [13]. In the dual situation (when $\omega^*(\Gamma) \subset K$ and $\omega(\Gamma)$ is a periodic orbit containing K in its interior), which could be discussed analogously using the reverse flow, it follows that K is a repeller with trivial shape. $\square$

From Theorem 3.1 it follows:

Corollary 3.1 *Let K be an isolated invariant continuum of a flow φ in $\mathbb{R}^2$. Suppose that K contains all the fixed points of φ and that there exists a bounded orbit γ in $\mathbb{R}^2 - K$. Then K is either an attractor or a repeller. Moreover, K has trivial shape.*

Proof. The set $K \cup \overline{\gamma}$ is compact and as a consequence there exists a closed disk D such that $K \cup \overline{\gamma} \subset \mathring{D}$. Then, Theorem 3.1 applies since the bounded orbit $\gamma \subset D - K$, and $D - K$ does not contain fixed points by assumption. □

Remark The assumptions about the existence of a disk D such that there is an entire orbit contained in $D - K$ in Theorem 3.1 and the existence of a bounded orbit in $\mathbb{R}^2 - K$ in Corollary 3.1 are unavoidable. For instance, consider the flow φ induced by the linear system

$$\begin{cases} \dot{x} = x \\ \dot{y} = -y \end{cases}$$

The origin $(0, 0)$ is a fixed point which is isolated as an invariant set and there are neither fixed points nor other bounded orbits in $\mathbb{R}^2 - \{(0, 0)\}$. In this case, $\{(0, 0)\}$ is a saddle and hence, it is neither an attractor nor a repeller.

As a consequence of Corollary 3.1 and [2, Theorem 12] we obtain the following dichotomy for dissipative flows:

Corollary 3.2 *Let K be an isolated invariant continuum of a dissipative flow φ in $\mathbb{R}^2$. Suppose that K contains all the fixed points of φ, then K has trivial shape and is either an attractor or a repeller. Moreover, if K is a repeller then there exists an attractor $K^* \subset \mathbb{R}^2 - K$ which is either a limit cycle or homeomorphic to a closed annulus bounded by two limit cycles.*

Proof. The dissipativeness of φ guarantees the existence of a global attractor K' and as a consequence $K \subset K'$. Suppose $K' \neq K$, since otherwise we have nothing to prove. Let $x \in K' - K$, the orbit $\gamma(x)$ is a bounded orbit being contained in the invariant compactum K'. Then, Corollary 3.1 ensures that K is either an attractor or a repeller. This proves the first part of the statement.

Suppose that K is a repeller and consider the flow $\varphi|K'$, i.e. the restriction of φ to the global attractor. The continuum K is also a repeller for $\varphi|K'$ and then there exists an invariant compactum $K^* \subset K'$ such that the pair (K^*, K) is an attractor-repeller decomposition of $\varphi|K'$. Besides, the invariant compactum K^* is an attractor for φ since K^* is an attractor for $\varphi|K'$ and K' is an attractor. The region of attraction of K^* agrees with $\mathbb{R}^2 - K$ since K is a repeller and (K^*, K) is an attractor-repeller decomposition of the restriction of φ to the global attractor K'. Moreover, $\mathbb{R}^2 - K$ is connected K being of trivial shape [4] and hence so is K^* by [13] and [4]. We have proved that K^* is a connected attractor which does not contain fixed points, thus by [2, Theorem 12] it must be either a limit cycle or homeomorphic to a closed annulus bounded by two limit cycles. □

References

[1] B. Alarcón, V. Guíñez, C. Gutiérrez: *Planar embeddings with a globally attracting fixed point*. Nonlinear Anal. **69** (2008) 140–150.

[2] H. Barge, J. M. R. Sanjurjo: *Unstable manifold, Conley index and fixed points of flows*. J. Math. Anal. Appl. **420** (2014) no. 1, 835–851.

[3] N. P. Bhatia, G. P. Szego: *Stability Theory of Dynamical Systems*. Grundlehren der Mat. Wiss. **16**, Springer, Berlin, 1970.

[4] K. Borsuk: *Theory of Shape*. Monografie Matematyczne **59**, Polish Scientific Publishers, Warsaw, 1975.

[5] C. Conley: *Isolated invariant sets and the Morse index*. CBMS Regional Conference Series in Mathematics **38** (Providence, RI: American Mathematical Society) 1978.

[6] C. Conley, R. W. Easton: *Isolated invariant sets and isolating blocks*. Trans. Amer. Math. Soc. **158** (1971) 35–61.

[7] R. W. Easton: *Isolating blocks and symbolic dynamics*. J. Diff. Equations **17** (1975) 96–118.

[8] I. Egawa: *A remark on the flow near a compact invariant set*. Proc. Japan Acad. **49** (1973) 247–251.

[9] B. M. Garay: *Uniform persistence and chain recurrence*. J. Math. Anal. Appl. **139** (1989) 372–381.

[10] C. Gutiérrez: *Smoothing continuous flows on two-manifolds and recurrences*. Ergod. Th. & Dynam. Sys. **6** (1986) 17–44.

[11] J. K. Hale: *Stability and gradient dynamical systems*. Rev. Mat. Complut. **17** (2004) 7–57.

[12] M. W. Hirsch, S. Smale, R. L. Devaney: *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Third Edition. Elsevier/Academic Press, Amsterdam, 2013.

[13] L. Kapitanski, I. Rodnianski: *Shape and Morse theory of attractors*. Comm. Pure Appl. Math. **53** (2000) 218–242.

[14] M. A. Morón, J. J. Sánchez Gabites, J. M. R. Sanjurjo: *Topology and dynamics of unstable attractors*. Fund. Math. **197** (2007) 239–252.

[15] J. Palis, W. de Melo: *Geometric Theory of Dynamical Systems. An introduction*. Springer-Verlag, New York-Berlin 1982.

[16] R. Ortega, F. R. Ruiz del Portal: *Attractors with vanishing rotation number*. J. Eur. Math. Soc. **13** (2011) 1569–1590.

[17] D. Salamon: *Connected simple systems and the Conley index of isolated invariant sets*. Trans. Amer. Math. Soc. **291** (1985) 1–41.

[18] J. J. Sánchez-Gabites: *Unstable attractors in manifolds*. Trans. Amer. Math. Soc. **362** (2010) 3563–3589.

[19] J. M. R. Sanjurjo: *On the fine structure of the global attractor of an uniformly persistent flow*. J. Diff. Equations **252** (2012) 4886–4897.

[20] T. Ura, I. Kimura: *Sur le courant exterieur a une region invariante. Théorème de Bendixson*. Comm. Mat. Univ. Sancii Pauli **8** (1960) 23–39.

The Contact Structure in the Space of Light Rays*

Alfredo BAUTISTA [(a)], **Alberto IBORT** [(a)] and **Javier LAFUENTE** [(b)]

(a) Departmento de Matemáticas
Universidad Carlos III de Madrid
28040 Madrid, Spain
abautist@math.uc3m.es albertoi@math.uc3m.es

(b) Departmento de Geometría y Topología
Universidad Complutense de Madrid
28040 Madrid, Spain
lafuente@mat.ucm.es

Dedicado a Jose María Montesinos con motivo de su jubilación.

ABSTRACT

The natural topological, differentiable and geometrical structures on the space of light rays of a given spacetime are discussed. The relation between the causality properties of the original spacetime and the natural structures on the space of light rays are stressed. Finally, a symplectic geometrical approach to the construction of the canonical contact structure on the space of light rays is offered.

2010 Mathematics Subject Classification: 53C50, 53D35, 58A30.
Key words: causal structure, strongly causal spacetime, null geodesic, light rays, contact structures.

1. Introduction

In the recent articles [Ba14, Ba15] it was shown that causality relations on spacetimes can be described alternatively in terms of the geometry and topology of the space of light rays and skies. This alternative description of causality, whose origin can be traced back to Penrose, was pushed forward by R. Low [Lo88, Lo06] and, as indicate above, largely accomplished in the referred works by Bautista, Ibort and Lafuente.

*This work has been partially supported by the Spanish MINECO project MTM2014-54692 and QUITEMAD+, S2013/ICE-2801.

Shifting the point of view from "events" to "light rays" and "skies" to analyze causality relations has deep and worth discussing implications. Thus, for instance, as Low himself noticed [Lo90, Lo94], in some instances, two events are causally related iff the corresponding skies are topologically linked and, a more precise statement of this fact, constitutes the so called Legendrian Low's conjecture (see for instance [Na04, Ch10]).

The existence of a canonical contact structure on the space of light rays plays a cornerstone role in this picture. Actually it was shown in [Ba15] that two events on a strongly causal spacetime are causally related iff there exists a non-negative sky Legendrian isotopy relating their corresponding skies.

Moreover, and as an extension of Penrose's twistor programme, it would be natural to describe attributes of the conformal class of the Lorentzian metric such as the Weyl tensor, in terms of geometrical structures on the space of light rays.

It is also worth to point out that in dimension 3 this dual approach to causality becomes very special. Actually if the dimension of spacetime is $m = 3$, the dimension of the space of light rays is also $2m - 3 = 3$ and the skies are just Legendrian circles. Even more because the dimension of the contact distribution is 2 the space of 1-dimensional subspaces is the 1-dimensional projective space $\mathbb{RP}^1$, hence the curve defined by the tangent spaces to the skies along the points of a light ray defines a projective segment of it and it is possible to define Low's causal boundary [Lo06] unambiguously. We will not dwell into these matters in the present paper and we will leave it to a detailed discussion elsewhere. (see also A. Bautista Ph. D. Thesis [Ba15b]).

Because of all these reasons we have found relevant to describe in a consistent and uniform way the fundamental structures present on the space of light rays of a given spacetime: that is, its topological, differentiable and geometrical structures, the later one, exemplified by its canonical contact structure. Thus, the present work will address in a sistematic and elementary way the description and construction of the aforementioned structures, highlighting the interplay between the causility properties of the original spacetime and the structures of the corresponding space of light rays.

The paper will be organized as follows. Section 2 will be devoted to introduce properly the space of light rays and in Sect. 3 its natural differentiable structure will be constructed. The tangent bundle of the space of light rays and the canonical contact structure on the space of light rays will be the subject of Sects. 4 and 5.

Looking for brevity in the present work, because a self–contained article will become a very long text, we suggest to the readers references [On83], [BE96], [HE73], [Mi08] and [Pe79] in order to review the basic elements of causality theory in Lorentzian manifolds.

2. The space $\mathcal{N}$ of light rays and its differentiable structure

2.1. Constructions of the space $\mathcal{N}$

Given a spacetime $(M, \mathbf{g})$, i.e., a Hausdoff, time-oriented, Lorentzian smooth manifold, we will denote by $\mathbb{N}$ the subset of all null vectors of $\widehat{T}M$, where in what follows, $\widehat{T}M$ will be used to make reference to the bundle resulting of eliminating the zero section of TM, that is $\widehat{T}M = \{v \in TM : v \neq 0\}$. The zero section of TM separates both connected components of $\mathbb{N}$ that will be denoted by

$$\mathbb{N}^+ = \{v \in \mathbb{N} : v \text{ future }\} \ , \qquad \mathbb{N}^- = \{v \in \mathbb{N} : v \text{ past }\} \ .$$

We will call the fibres $\mathbb{N}_p$, $\mathbb{N}_p^+$ and $\mathbb{N}_p^-$ *lightcone*, *future lightcone* and *past lightcone* at $p \in M$ respectively. We define the *set of light rays* of $(M, \mathbf{g})$ by

$$\mathcal{N}_{\mathbf{g}} = \{\text{Im } (\gamma) \subset M : \gamma \text{ is a maximal null geodesic in } (M, \mathbf{g})\}$$

where Im (γ) denotes the image of the curve γ. This definition seems to depend on the metric $\mathbf{g}$, however we will show that the space of light rays depends only on the conformal class of the spacetime metric.

We define the *conformal class of metrics in M equivalent to* $\mathbf{g}$ by

$$\mathcal{C}_{\mathbf{g}} = \left\{\overline{\mathbf{g}} \in \mathfrak{T}_0^2 (M) : \overline{\mathbf{g}} = e^{2\sigma}\mathbf{g}, \ \ 0 < \sigma \in \mathfrak{F}(M)\right\}$$

and we call $(M, \mathcal{C}_{\mathbf{g}})$ the corresponding *conformal class of spacetimes equivalent to* $(M, \mathbf{g})$.

It is known that two metric are conformally equivalent if the lightcones coincide at every point (see [Mi08, Prop. 2.6 and Lem. 2.7] or [HE73, p. 60-61]). This fact can be automatically translated to the spaces of light rays defined by two different metrics on a manifold M. The following proposition brings to light the equivalence among spaces of light rays.

Proposition 2.1 *Let $(M, \mathbf{g})$ and $(M, \overline{\mathbf{g}})$ be two spacetimes and let $\mathcal{N}_{\mathbf{g}}$ and $\mathcal{N}_{\overline{\mathbf{g}}}$ be their corresponding spaces of light rays. Then $(M, \mathbf{g})$ and $(M, \overline{\mathbf{g}})$ are conformally equivalent if and only if $\mathcal{N}_{\mathbf{g}} = \mathcal{N}_{\overline{\mathbf{g}}}$.*

Because of Prop, 2.1 we have the following definition.

Definition 2.1 *Let $(M, \mathcal{C}_{\mathbf{g}})$ be a conformal class of spacetimes with* $\dim M = m \geq 3$. *We will call* light ray *the image $\gamma(I)$ in M of a maximal null geodesic $\gamma : I \to M$ related to any metric $\bar{\mathbf{g}} \in \mathcal{C}_{\mathbf{g}}(M)$. It will be denoted by $[\gamma]$ or γ when there is not possibility of confusion, that is $[\gamma] \in \mathcal{N}$, $\gamma \in \mathcal{N}$ or also $\gamma \subset M$. So, every light ray is equivalent to an unparametrized null geodesic. Then, we will say that the* space of light rays $\mathcal{N}$ of a conformal class of spacetimes $(M, \mathcal{C}_{\mathbf{g}})$ *is the set*

$$\mathcal{N} = \{\gamma(I) \subset M \mid \gamma : I \to M \text{ is a maximal null geodesic for any metric } \bar{\mathbf{g}} \in \mathcal{C}_{\mathbf{g}}\} \ .$$

A more geometric construction of $\mathcal{N}$ is possible as a quotient space of the tangent bundle TM [Lo06]. This construction will allow us to show how $\mathcal{N}$ inherits the topological and differentiable structures of TM.

Let us consider the *geodesic spray* $X_{\mathbf{g}}$ related to the metric $\mathbf{g}$, that is the vector field in TM such that its integral curves define the geodesics in $(M, \mathbf{g})$ and their tangent vectors, and the *Euler field* Δ *in* TM, the vector field in TM whose flow are scale transformations along the fibres of TM.

It is easy to see that both $X_{\mathbf{g}}$ and Δ are tangent to $\mathbb{N}$, and the differentiable distribution in $\mathbb{N}^+$ given by $\mathcal{D} = \operatorname{span}\{X_{\mathbf{g}}, \Delta\}$ verifies

$$[\Delta, X_{\mathbf{g}}] = X_{\mathbf{g}}\,,$$

hence $\mathcal{D}$ is involutive and, by Fröbenius' Theorem [Wa83, Thm. 1.60], it is also integrable. This means that the quotient space $\mathbb{N}^+/\mathcal{D}$ is well defined. Every leaf of $\mathcal{D}$ is the equivalence class consisting of a future–directed null geodesic and all its affine reparametrizations preserving time–orientation, hence the space

$$\mathcal{N}^+ = \mathbb{N}^+/\mathcal{D}\,.$$

is the space of future-oriented light rays of M. In a similar way we may construct the space of past-oriented light rays $\mathcal{N}^- = \mathbb{N}^-/\mathcal{D}$.

The space of light rays $\mathcal{N}$ can be obtained as the quotient $\mathbb{N}/\widetilde{\mathcal{D}}$ where $\widetilde{D}$ denotes the scale transformation group acting on $\mathbb{N}$, that is, $v \mapsto \lambda v$, $\lambda \neq 0$, $v \in \mathbb{N}$. Notice that $\widetilde{\mathcal{D}} \cong \mathbb{R} - \{0\}$ and the orbits of the connected component containing the identity can be identified with the leaves of $\mathcal{D}$.

Because our spacetime M is time-orientable, in what follows we will consider the space of future-oriented light rays $\mathcal{N}^+$ and we will denote it again as $\mathcal{N}$ without risk of confusion. As it will be shown later on, this convention will be handy as the space $\mathcal{N}^+$ carries a co-orientable contact structure (see Sect. 4) whereas $\mathcal{N}$ does not.

2.2. Differentiable structure of $\mathcal{N}$

If we require the space of light rays of M to be a differentiable manifold, it is necessary to ensure that the leaves of the distribution that builds $\mathcal{N}$, are regular submanifolds. This is not automatically true for any spacetime M, as example [Lo01, Ex. 1] shows, so it will be necessary to impose further conditions to ensure it.

It important to observe that if M is strongly causal, every causal curve has a finite number of connected components in a relatively compact neighbourhood, and therefore the distribution $\mathcal{D} = \operatorname{span}\{X_{\mathbf{g}}, \Delta\}$ is regular, then the quotient $\mathcal{N} = \mathbb{N}^+/\mathcal{D}$ is a manifold, as the following proposition [Lo89, Pr. 2.1] claims.

Proposition 2.2 *Let M be a strongly causal spacetime, then the distribution above defined by $\mathcal{D} = \operatorname{span}\{X_{\mathbf{g}}, \Delta\}$ is regular and the space of light rays $\mathcal{N}$ inherits from $\mathbb{N}^+$ the structure of differentiable manifold such that $p_{\mathbb{N}^+} : \mathbb{N}^+ \to \mathcal{N}$ defined by $p_{\mathbb{N}^+}(u) = [\gamma_u]$ is a submersion.*

The space $\mathcal{N}$ can also be constructed as a quotient of the bundle of null directions $\mathbb{PN}$ defined below, that is we can proceed to compute the quotient space $\mathbb{N}^+/\mathcal{D}$ in two steps, we will first quotient with respect to the action dilation field Δ and, secondly we will pass to the quotient defined by the integral curves of the geodesic field.

First, consider the involutive distribution $\mathcal{D}_\Delta = \operatorname{span}\{\Delta\}$. Observe that for any $u \in TM$ the vectors $e^t u \in TM$ run the integral curve of Δ passing through u when $t \in \mathbb{R}$. Then, if

$$\phi\colon TV \to \mathbb{R}^{2m}; \quad \xi \mapsto (x, \dot{x}) \tag{2.1}$$

are the canonical coordinates in TM, then $(x, e^t \dot{x})$ describe integral curves of Δ. Taking homogeneous coordinates for $\dot{x}$, we can obtain that $(x, [\dot{x}])$ are coordinates adapted to foliation generated by $\mathcal{D}_\Delta$. Therefore $\mathcal{D}_\Delta$ is regular and its restriction to $\mathbb{N}^+$ is also regular, hence the quotient space $\mathbb{N}^+/\mathcal{D}_\Delta$ defined by

$$\mathbb{PN} = \mathbb{N}^+/\mathcal{D}_\Delta = \left\{[\xi] : \eta \in [\xi] \Leftrightarrow \eta = e^t \xi \text{ for some } t \in \mathbb{R} \text{ and } \xi \in \mathbb{N}^+\right\}$$

is a differentiable manifold and, moreover, the canonical projection $\pi^{\mathbb{N}^+}_{\mathbb{PN}} : \mathbb{N}^+ \to \mathbb{PN}$, given by $\xi \mapsto [\xi]$, is a submersion.

The next step is to find a regular distribution that allows us to define $\mathcal{N}$ by a quotient. For each vector $u \in \mathbb{N}^+_p$ there exists a null geodesic γ_u such that $\gamma_u(0) = p$ and $\gamma'_u(0) = u$, and given two vectors $u, v \in \mathbb{N}^+_p$ verifying that $v = \lambda u$ with $\lambda > 0$, then the geodesics γ_u and γ_v such that $\gamma_u(0) = \gamma_v(0) = p$ have the property

$$\gamma_v(s) = \gamma_{\lambda u}(s) = \gamma_u(\lambda s)$$

hence they have the same image in M and then $\gamma_v = \gamma_u$ as unparametrized sets in M. This fact implies that the elevations to $\mathbb{PN}$ of the null geodesics of M define a foliation $\mathcal{D}_G$. Two directions $[u], [v] \in \mathbb{PN}$ belong to the same leaf of the foliation $\mathcal{D}_G$ if for the vectors $v \in \mathbb{N}^+_p$ and $u \in \mathbb{N}^+_q$ there exist null geodesics γ_1 and γ_2 and values $t_1, t_2 \in \mathbb{R}$ verifying

$$\left\{\begin{array}{l} \gamma_1(t_1) = p \in M \\ \gamma'_1(t_1) = v \in \mathbb{N}^+_p \end{array}\right. \qquad \text{and} \qquad \left\{\begin{array}{l} \gamma_2(t_2) = q \in M \\ \gamma'_2(t_2) = u \in \mathbb{N}^+_q \end{array}\right.$$

such that there is a reparametrization h verifying $\gamma_1 = \gamma_2 \circ h$.

Hence, the space of leaves of $\mathcal{D}_G$ in $\mathbb{PN}$ coincides with $\mathcal{N}$, that is,

$$\mathcal{N} = \mathbb{PN}/\mathcal{D}_G$$

The map $p_{\mathbb{PN}} : \mathbb{PN} \longrightarrow \mathcal{N}$ given by $[u] \longmapsto [\gamma_u]$ is well defined, since $\gamma_{\lambda u}(s) = \gamma_u(\lambda s)$ as seen above, and it verifies the identity $p_{\mathbb{PN}}([\gamma'_u(s)]) = [\gamma_u] \in \mathcal{N}$ for all s.

Proposition 2.2 can be formulated for the bundle $\mathbb{PN}$ with regular distribution$\mathcal{D}_G$ instead. In [Lo06, Thm. 1] a similar result is proved for the subbundle $(\mathbb{N}^*)^+$ of the cotangent bundle T^*M.

Now, we will describe a generic way to construct coordinate charts in $\mathcal{N}$. First, we define for any subset $W \subset M$, the sets $\mathbb{N}(W) = \left\{ \xi \in \mathbb{N} : \pi_M^{\mathbb{N}}(\xi) \in W \subset M \right\}$, and $\mathbb{PN}(W) = \left\{ [\xi] \in \mathbb{PN} : \pi_M^{\mathbb{PN}}([\xi]) \in W \subset M \right\}$.

Take $V \subset M$ a causally convex, globally hyperbolic and open set with a smooth spacelike Cauchy surface $C \subset V$. Let $\mathcal{U}$ be the image of the projection $p_{\mathbb{N}} : \mathbb{N}^+(V) \mapsto \mathcal{N}$. Since $\mathbb{N}^+(V)$ is open in $\mathbb{N}^+$ and $p_{\mathbb{N}}$ is a submersion, then $\mathcal{U} \subset \mathcal{N}$ is open. Each null geodesic passing through V intersects C in a unique point and since $p_{\mathbb{N}^+} = p_{\mathbb{PN}} \circ \pi_{\mathbb{PN}}^{\mathbb{N}^+}$, this ensures that

$$\mathcal{U} = p_{\mathbb{N}}\left(\mathbb{N}^+(V)\right) = p_{\mathbb{N}}\left(\mathbb{N}^+(C)\right) = p_{\mathbb{N}} \circ \pi_{\mathbb{PN}}^{\mathbb{N}^+}\left(\mathbb{N}^+(C)\right) = p_{\mathbb{PN}}\left(\mathbb{PN}(C)\right) = p_{\mathbb{PN}}\left(\mathbb{PN}(V)\right).$$

Then we can define the diffeomorphism $\sigma = p_{\mathbb{PN}}|_{\mathbb{PN}(C)} : \mathbb{PN}(C) \mapsto \mathcal{U}$. So, we have the following diagram

$$\begin{tikzcd} \mathbb{PN}(V) \arrow[r, "p_{\mathbb{PN}}"] & \mathcal{U} \\ \mathbb{PN}(C) \arrow[u, hook, "\mathrm{inc}"] \arrow[ur, "\sigma"'] & \end{tikzcd} \tag{2.2}$$

If ϕ is any coordinate chart for $\mathbb{PN}(C)$ then $\phi \circ \sigma^{-1}$ is a coordinate chart for $\mathcal{U} \subset \mathcal{N}$.

Observe that if M is time–orientable, there exists a non–vanishing future timelike vector field $T \in \mathfrak{X}(M)$ everywhere. Then we can define the submanifold $\Omega^T(C) \subset \mathbb{N}^+(C)$ by

$$\Omega^T(C) = \left\{ \xi \in \mathbb{N}^+(C) : \mathbf{g}(\xi, T) = -1 \right\}$$

The restriction $\pi_{\mathbb{PN}}^{\mathbb{N}^+}\Big|_{\Omega^T(C)} : \Omega^T(C) \to \mathbb{PN}(C)$ of the submersion $\pi_{\mathbb{PN}}^{\mathbb{N}^+} : \mathbb{N}^+ \to \mathbb{PN}$ is clearly a diffeomorphism.

So, we have the following diagram

$$\mathcal{N} \supset \mathcal{U} \leftrightarrow \mathbb{PN}(C) \leftrightarrow \Omega^T(C) \hookrightarrow \mathbb{N}^+(C) \hookrightarrow \mathbb{N}^+ \hookrightarrow TM \tag{2.3}$$

where $\leftrightarrow$ and $\hookrightarrow$ represent diffeomorphisms and inclusions respectively.

Then, the composition of the diffeomorphism $\mathcal{U} \to \Omega^T(C)$ with the restriction of a coordinate chart in TM to the vectors in $\Omega^T(C)$, can be used to construct a coordinate chart in $\mathcal{N}$.

If we require the space of light rays of M to be a differentiable manifold, it remains to ensure that $\mathcal{N}$ is a Hausdorff topological space. Again, it is not verified for any strongly causal spacetime M as we can check in example 2.3, so we need to state conditions to ensure it.

Example 2.3 *$\mathcal{N}$ is not Hausdorff. Consider the two–dimensional Minkowski spacetime and remove the point $(1,1)$. Clearly, M is strongly causal. Let $\{\tau_n\} \subset \mathbb{R}$ be a sequence such that $\lim_{n\mapsto\infty} \tau_n = 0$. Then the sequence of null geodesic given by $\lambda_n(s) = (s, \tau_n + s)$ with $s \in (-\infty, \infty)$ converges to two different null geodesics, $\mu_1(s) = (s,s)$ with $s \in (-\infty, 1)$ and $\mu_2(s) = (s,s)$ with $s \in (1, \infty)$.*

A sufficient condition to ensure that $\mathcal{N}$ is Hasudorff is the absence of naked singularities [Lo89, Prop. 2.2], but we will see in example 2.4 that it is not a necessary condition.

Example 2.4 *Let $\mathbb{M}$ be the 3–dimensional Minkowski space–time described by coordinates (t,x,y) and equipped with the metric $\mathbf{g} = -dt \otimes dt + dx \otimes dx + dy \otimes dy$. Now, consider the restriction $\mathbb{B} = \left\{(t,x,y) \in \mathbb{M} : t^2 + x^2 + y^2 < 1\right\}$. It is clear that $\mathbb{B}$ is strongly causal but not globally hyperbolic.*

It is easy to find inextensible causal curves fully contained in the chronological future or past of some event $p \in \mathbb{B}$, then $\mathbb{B}$ is nakedly singular.

The space of light rays $\mathcal{N}_{\mathbb{B}}$ of $\mathbb{B}$ verifies that $\mathcal{N}_{\mathbb{B}} \subset \mathcal{N}_{\mathbb{M}}$. Moreover, the inclusion $\mathcal{N}_{\mathbb{B}} \hookrightarrow \mathcal{N}_{\mathbb{M}}$ is clearly injective and every light ray has a neighbourhood in $\mathcal{N}_{\mathbb{B}}$ corresponding to another neighbourhood in $\mathcal{N}_{\mathbb{M}}$. It is not difficult to show that $\mathcal{N}_{\mathbb{B}}$ is open in $\mathcal{N}_{\mathbb{M}}$. Since M is globally hyperbolic, then $\mathcal{N}_{\mathbb{M}}$ is Hausdorff and therefore $\mathcal{N}_{\mathbb{B}}$ is also Hausdorff.

Example 2.4 shows that the absence of naked singularities is a condition too strong for a strongly causal spacetime M. Moreover in this case, M becomes globally hyperbolic as Penrose proved in [Pe79].

A suitable condition to avoid the behavior of light rays in the paradigmatic example 2.3 but to permit naked singularities similar to the ones in example 2.4 is the condition of *null pseudo–convexity.*

Definition 2.2 *A spacetime M is said to be* null pseudo–convex *if for any compact $K \subset M$ there exists a compact $K' \subset M$ such that any null geodesic segment γ with endpoints in K is totally contained in K'.*

In [Lo90-2], Low states the equivalence of null pseudo–convexity of M and the Hausdorffness of $\mathcal{N}$ for a strongly causal spacetime M. From now on, we will assume that M is a strongly causal and null pseudo–convex spacetime unless others conditions are pointed out.

3. Tangent bundle of $\mathcal{N}$

To take advantage of the geometry and topology of $\mathcal{N}$ it is needed to have a suitable characterization of the tangent spaces $T_\gamma\mathcal{N}$ for any $\gamma \in \mathcal{N}$. We will proceed as follows: first, we fix an auxiliary metric in the conformal class $\mathcal{C}$, then we will define *geodesic variations* (in particular, *variations by light rays*) and *Jacobi fields*, explaining the relation between both concepts (in lemmas 3.1, 3.2, 3.3 and proposition 3.2). Then, in proposition 3.3, we will characterize tangent vector of TM in terms of Jacobi fields. Second, we will keep an eye on how the Jacobi fields changes when we vary the parameters of the corresponding variation by light rays or conformal metric of M (see from lemma 3.4 to 3.9).

Finally, in proposition 3.4, we will get the aim of this section identifying tangent vectors of $\mathcal{N}$ with some equivalence classes of Jacobi fields.

Definition 3.1 *A differentiable map* $\mathbf{x} : (a,b) \times (\alpha,\beta) \to M$ *is said to be a* variation of a segment of curve $c : (\alpha,\beta) \to M$ *if* $c(t) = \mathbf{x}(s_0,t)$ *for some* $s_0 \in (a,b)$. *We will say that* $V^{\mathbf{x}}_{s_0}$ *is the* initial field of $\mathbf{x}$ *in* $s = s_0$ *if*

$$V^{\mathbf{x}}_{s_0}(t) = d\mathbf{x}_{(s_0,t)}\left(\frac{\partial}{\partial s}\right)_{(s_0,t)} = \left.\frac{\partial \mathbf{x}(s,t)}{\partial s}\right|_{(s_0,t)}$$

defining a vector field along c.

We will say that $\mathbf{x}$ *is a* geodesic variation *if any longitudinal curve of* $\mathbf{x}$, *that is* $c^{\mathbf{x}}_s = \mathbf{x}(s,\cdot)$ *for* $s \in (a,b)$, *is a geodesic.*

If the longitudinal curves $c^{\mathbf{x}}_s : (\alpha,\beta) \to M$ *are regular curves covering segments of light rays, then* $\mathbf{x} : (a,b) \times (\alpha,\beta) \to M$ *is said to be a* variation by light rays.

Moreover, a variation by light rays $\mathbf{x}$ *is said to be a* variation by light rays of $\gamma \in \mathcal{N}$ *if* γ *is a longitudinal curve of* $\mathbf{x}$.

Notation 3.1 *It is possible to identify a given segment of null geodesic* $\gamma : (-\delta,\delta) \to M$, *with a slight abuse in the notation, to the light ray in* $\mathcal{N}$ *defined by it. So, if* $\mathbf{x} = \mathbf{x}(s,t)$ *is a variation by light rays, we can denote by* $\gamma^{\mathbf{x}}_s \subset M$ *the null pregeodesics of the variation and also by* $\gamma^{\mathbf{x}}_s \in \mathcal{N}$ *the light rays they define.*

Consider a geodesic curve $\mu(t)$ in a spacetime $(M,\mathbf{g})$. Given $J \in \mathfrak{X}_\mu$, we will abbreviate the notation $J' = \frac{DJ}{dt}$ and $J'' = \frac{D}{dt}\frac{DJ}{dt} = \frac{D^2 J}{dt^2}$. We can define the *Jacobi equation* by

$$J'' + R(J,\mu')\mu' = 0 \tag{3.1}$$

where R is the Riemann tensor. We will name the solutions of the equation 3.1 by *Jacobi field* along μ. So, the set of Jacobi fields along μ is then defined by

$$\mathcal{J}(\mu) = \{J \in \mathfrak{X}_\mu : J'' + R(J,\mu')\mu' = 0\} \tag{3.2}$$

The linearity of $\frac{D}{dt}$ and R provides a vector space structure to $\mathcal{J}(\mu)$ hence $\mathcal{J}(\mu)$ is a vector subspace of $\mathfrak{X}_\mu$.

The relation between geodesic variations and Jacobi fields is expounded in next lemma.

Lemma 3.1 *If* $\mathbf{x} : (-\epsilon,\epsilon) \times (-\delta,\delta) \to M$ *is a geodesic variation of a geodesic* γ, *then the initial field* $V^{\mathbf{x}}$ *is a Jacobi field along* γ.

Proof. See [On83, Lem. 8.3]. □

A Jacobi field along a geodesic γ is fully defined by its initial values at any point of γ as lemma 3.2 claims, and moreover it also implies that the vector space $\mathcal{J}(\mu)$ is isomorphic to $T_pM \times T_pM$ therefore $\dim(\mathcal{J}(\gamma)) = 2\dim(M) = 2m$.

Lemma 3.2 *Let γ be a geodesic in M such that $\gamma(0) = p$ and $u, v \in T_pM$. Then there exists a only Jacobi field J along γ such that $J(0) = u$ and $\frac{DJ}{dt}(0) = v$.*

Proof. See [On83, Lem. 8.5]. □

Next lemma characterizes the Jacobi fields of a particular type of variation. This type will be the general case for the variations by light rays studied below.

Lemma 3.3 *Let M be a spacetime, $\gamma : (-\delta, \delta) \to M$ a geodesic segment, $\lambda : (-\epsilon, \epsilon) \to M$ a curve verifying $\lambda(0) = \gamma(0)$, and $W(s)$ a vector field along λ such that $W(0) = \gamma'(0)$. Then the Jacobi field J along γ defined by the geodesic variation*

$$\mathbf{x}(s,t) = \exp_{\lambda(s)}(tW(s))$$

verifies that

$$J(0) = \lambda'(0), \qquad \frac{DJ}{dt}(0) = \frac{DW}{ds}(0)$$

Proof. First, the vector $\frac{\partial \mathbf{x}}{\partial s}(0,0)$ is the tangent vector of the curve $\mathbf{x}(s,0)$ in $s = 0$, and since $\mathbf{x}(s,0) = \exp_{\lambda(s)}(0 \cdot W(s)) = \exp_{\lambda(s)}(0) = \lambda(s)$, then we have

$$J(0) = \frac{\partial \mathbf{x}}{\partial s}(0,0) = \frac{d\lambda}{ds}(0) = \lambda'(0)$$

On the other hand, $\frac{D}{ds}\frac{\partial \mathbf{x}}{\partial t}(0,0)$ is the covariant derivative of the vector field $\frac{\partial \mathbf{x}}{\partial t}(s,0) = W(s)$ for $s = 0$ along the curve $\mathbf{x}(s,0) = \lambda(s)$. Then

$$\frac{DJ}{dt}(0) = \frac{D}{dt}\frac{\partial \mathbf{x}}{\partial s}(0,0) = \frac{D}{ds}\frac{\partial \mathbf{x}}{\partial t}(0,0) = \frac{DW}{ds}(0).$$

as required. □

Remark 3.1 *It can be observed that given a geodesic variation $\mathbf{x} = \mathbf{x}(s,t)$ such that J is the corresponding Jacobi field at $s = 0$, if we change the geodesic parameters such that $\overline{\mathbf{x}}(s,\tau) = \mathbf{x}(s, a\tau + b)$ for $a > 0$ and $b \in \mathbb{R}$, then the Jacobi field $\overline{J}$ of $\overline{\mathbf{x}}$ at $s = 0$ verify $\overline{J} = J$. This implies that changing the geodesic parameter does not modify the Jacobi field as a geometric object.*

Proposition 3.2 *Given a geodesic γ in $(M, \mathbf{g})$ and a Jacobi field $J \in \mathcal{J}(\gamma)$ along γ, then $\mathbf{g}(J(t), \gamma'(t)) = a + bt$ is verified.*

Proof. It is trivial deriving $\mathbf{g}(J(t), \gamma'(t))$ twice with respect to t. □

Remark 3.2 *Observe the following fact that we will need later. Given a differentiable curve $v : I \to TM$, the information contained in the tangent vector $v'(s_0) \in TTM$ coincides with the one in the covariant derivative of v as vector field along its base curve in M. That is, from $v'(s_0)$ it is possible to determine the vectors $\alpha'(s_0)$ and the covariant derivative $\frac{Dv}{ds}(s_0)$ along α, where $\alpha = \pi_M^{TM} \circ v$, and vice–versa.*

It is possible to identify any tangent vector $\xi \in TTM$ with a Jacobi field along the geodesic γ defined by the exponential of the vector $u = \pi_{TM}^{TTM}(\xi) \in TM$.

Proposition 3.3 *Given a vector $u_0 \in T_pM$ and consider the geodesic γ_{u_0} defined by $\gamma_{u_0}(t) = \exp_p(tu_0)$. Let $u : (-\delta, \delta) \to TM$ be a differentiable curve such that $u(0) = u_0$ and $u'(0) = \xi$. If $J \in \mathcal{J}(\gamma_{u_0})$ is the Jacobi field of the geodesic variation given by $\mathbf{x}(s,t) = \exp_{\alpha(s)}(tu(s))$ where $\alpha = \pi_M^{TM} \circ u$, then the map*

$$\begin{array}{rccc} \zeta : & T_{u_0}TM & \to & \mathcal{J}(\gamma_{u_0}) \\ & \xi & \mapsto & J \end{array}$$

is a well–defined linear isomorphism.

Proof. Let $u_i : (-\delta_i, \delta_i) \to TM$ be differentiable curves such that $u_i(0) = u_0$ and $u_i'(0) = \xi$ for $i = 1, 2$, and consider the geodesic variations $\mathbf{x}_i(s_i, t) = \exp_{\alpha(s_i)}(tu_i(s_i))$. First, observe that for every $\xi \in T_{u_0}TM$, remark 3.2 and lemma 3.3 imply that ξ defines unambiguously the initial values of the Jacobi field of $\mathbf{x}$ at $s = 0$. So ζ is well–defined.

The linearity of ζ is straightforward using remark 3.2.

Finally, if $\xi \in TM$ such that $\zeta(\xi) = 0$ then, in virtue of lemmas 3.3 and remark 3.2, we have that $\xi = 0$. This implies that ζ is injective, but since $\dim(\mathcal{J}(\gamma_{u_0})) = \dim(T_{u_0}TM) = 2m$ then we conclude that ζ is an isomorphism. □

Now, we will focus on the variations by light rays and the Jacobi fields they define. Fix a null geodesic $\gamma \in \mathcal{N}$ and assume that $\mathbf{x}(s,t)$ is a variation by light rays of $\gamma = \gamma_0^{\mathbf{x}} \in \mathcal{N}$ in such a way that $J(t) = V_0^{\mathbf{x}}(t)$ is the Jacobi field over γ corresponding to the initial field of $\mathbf{x}$ and $\frac{\partial \mathbf{x}}{dt}(s,t) = (\gamma_s^{\mathbf{x}})'(t)$. Since $\mathbf{x}$ is a variation by light rays, then it provides that $\mathbf{g}\left((\gamma_s^{\mathbf{x}})'(t), (\gamma_s^{\mathbf{x}})'(t)\right) = 0$ for all (s,t) in the domain of $\mathbf{x}$, hence

$$0 = \left.\frac{\partial}{\partial s}\right|_{(0,t)} \mathbf{g}\left((\gamma_s^{\mathbf{x}})'(t), (\gamma_s^{\mathbf{x}})'(t)\right) = 2\mathbf{g}\left(\left.\frac{D}{ds}\right|_{(0,t)} \frac{\partial \mathbf{x}}{dt}(s,t), \frac{\partial \mathbf{x}}{dt}(0,t)\right) =$$

$$= 2\mathbf{g}\left(\left.\frac{D}{dt}\right|_{(0,t)} \frac{\partial \mathbf{x}}{ds}(s,t), \frac{\partial \mathbf{x}}{dt}(0,t)\right) = \left.\frac{\partial}{\partial t}\right|_{(0,t)} \mathbf{g}\left(V_s^{\mathbf{x}}(t), (\gamma_s^{\mathbf{x}})'(t)\right) =$$

$$= \left.\frac{\partial}{\partial t}\right|_{t} \mathbf{g}\left(V^{\mathbf{x}}(t), \gamma'(t)\right) = \left.\frac{\partial}{\partial t}\right|_{t} \mathbf{g}\left(J(t), \gamma'(t)\right)$$

then the variations by light rays of a null geodesic γ verify that their Jacobi fields J fulfil

$$\mathbf{g}\left(J\left(t\right),\gamma'\left(t\right)\right)=c$$

with $c\in\mathbb{R}$ constant. Then, we define the set of Jacobi fields of variations by light rays by

$$\mathcal{J}_L\left(\gamma\right)=\left\{J\in\mathcal{J}\left(\gamma\right):\mathbf{g}\left(J,\gamma'\right)=c\text{ constant}\right\}$$

Since $\mathbf{g}$ is bilinear then $\mathcal{J}_L\left(\gamma\right)$ is a vector subspace of $\mathcal{J}\left(\gamma\right)$ verifying $\dim\left(\mathcal{J}_L\left(\gamma\right)\right)=2\dim\left(M\right)-1=2m-1$.

Now, we define one–dimensional subspaces of $\mathcal{J}\left(\gamma\right)$ given by

$$\widehat{\mathcal{J}}_0\left(\gamma\right)=\left\{J\left(t\right)=bt\gamma'\left(t\right):b\in\mathbb{R}\right\}\ ,\qquad \widehat{\mathcal{J}}_0'\left(\gamma\right)=\left\{J\left(t\right)=a\gamma'\left(t\right):a\in\mathbb{R}\right\}\ .$$

It is trivial to see that $\widehat{\mathcal{J}}_0\left(\gamma\right)\subset\mathcal{J}_L\left(\gamma\right)$ and $\widehat{\mathcal{J}}_0'\left(\gamma\right)\subset\mathcal{J}_L\left(\gamma\right)$.

If $J\in\widehat{\mathcal{J}}_0\left(\gamma\right)\cap\widehat{\mathcal{J}}_0'\left(\gamma\right)$, then its initial values must verify

$$\left\{\begin{array}{l}J\left(0\right)=0\\ J'\left(0\right)=b\gamma'\left(0\right)\end{array}\right.\qquad\text{and}\qquad\left\{\begin{array}{l}J\left(0\right)=a\gamma'\left(0\right)\\ J'\left(0\right)=0\end{array}\right.$$

then $a=b=0$ and therefore $\widehat{\mathcal{J}}_0\left(\gamma\right)\cap\widehat{\mathcal{J}}_0'\left(\gamma\right)=\{0\}$. So, we can define the direct product

$$\mathcal{J}_0\left(\gamma\right)=\widehat{\mathcal{J}}_0\left(\gamma\right)\oplus\widehat{\mathcal{J}}_0'\left(\gamma\right)=\left\{J\left(t\right)=\left(a+bt\right)\gamma'\left(t\right):a,b\in\mathbb{R}\right\}$$

being the vector subspace of Jacobi fields proportional to γ' and verifying $\dim\left(\mathcal{J}_0\left(\gamma\right)\right)=2$.

Now, we can define the quotient vector space

$$\mathcal{L}\left(\gamma\right)=\mathcal{J}_L\left(\gamma\right)/\mathcal{J}_0\left(\gamma\right)=\left\{[J]:K\in[J]\Leftrightarrow K=J+J_0\text{ such that }J_0\in\mathcal{J}_0\left(\gamma\right)\right\}$$

whose dimension is $\dim\left(\mathcal{L}\left(\gamma\right)\right)=\dim\left(\mathcal{J}_L\left(\gamma\right)\right)-\dim\left(\mathcal{J}_0\left(\gamma\right)\right)=2\dim\left(M\right)-3$. The elements of $\mathcal{L}\left(\gamma\right)$ will be denoted by $[J]\equiv J\left(\mathrm{mod}\gamma'\right)$ and we will say that $K=J\left(\mathrm{mod}\gamma'\right)$ when $[K]=[J]$.

Next lemma claims that there exist a change of parameter such that any variation by light rays can be transformed in a geodesic variation by light rays. So, lemma 3.1 can be used.

Lemma 3.4 *Let $\mathbf{x}=\mathbf{x}\left(s,t\right)$ be a variation by light rays in $\left(M,\mathcal{C}\right)$ such that $\gamma_s\left(t\right)=\mathbf{x}\left(s,t\right)$ defines its light rays. Fixed any metric $\mathbf{g}\in\mathcal{C}$ then there exists a differentiable function $h=h\left(s,\tau\right)$ such that the light rays parametrized as $\overline{\gamma}_s=\gamma_s\left(h\left(s,\tau\right)\right)$ are null geodesics related to $\mathbf{g}$.*

Proof. Since each γ_s is a segment of light ray then $\gamma_s=\gamma_s\left(t\right)$ is a pregeodesic related to $\mathbf{g}$. Hence

$$\frac{D\gamma_s'\left(t\right)}{dt}=\frac{D}{dt}\frac{\partial\mathbf{x}}{\partial t}\left(s,t\right)=f\left(s,t\right)\gamma_s'\left(t\right)$$

where f is differentiable and $\frac{D}{dt}$ denotes the covariant derivative related to $\mathbf{g}$ along $\gamma_s(t)$.

It is enough to check that the function $h(s,\tau) = h_s(\tau)$ defined by

$$h_s^{-1}(t) = \int_0^t e^{\int_0^x f(s,y)dy} dx \tag{3.3}$$

makes of $\overline{\gamma}_s = \gamma_s \circ h$ a null geodesic with respect to the metric $\mathbf{g}$. □

Lemma 3.5 shows that any differentiable curve $\Gamma \subset \mathcal{N}$ defines a variation by light rays $\mathbf{x}$ such that the longitudinal curves of $\mathbf{x}$ corresponds to points in Γ. This variation is not unique by construction.

Lemma 3.5 *Given a differentiable curve $\Gamma : I \to \mathcal{N}$ such that $0 \in I$ and $\Gamma(s) = \gamma_s \subset M$, then there exists a variation by light rays $\mathbf{x} : (-\epsilon,\epsilon) \times (-\delta,\delta) \to M$ verifying*

$$\mathbf{x}(s,t) = \gamma_s(t)$$

for all $(s,t) \in (-\epsilon,\epsilon) \times (-\delta,\delta)$. Moreover, the variation $\mathbf{x}$ can be written as

$$\mathbf{x}(s,t) = \exp_{\pi_M^{\mathbb{N}^+}(v(s))}(tv(s))$$

where $v : (-\epsilon,\epsilon) \to \mathbb{N}^+(C)$ is a differentiable curve.

Proof. Consider the restriction $\pi = \left.\pi_{\mathbb{PN}}^{\mathbb{N}^+}\right|_{\mathbb{N}^+(C)} : \mathbb{N}^+(C) \to \mathbb{PN}(C)$ and the diffeomorphism $\sigma : \mathbb{PN}(C) \to \mathcal{U}$ in the diagram 2.2, where $\mathcal{U} \subset \mathcal{N}$ and $V \subset M$ are open such that V is globally hyperbolic and $C \subset V$ is a Cauchy surface of V and moreover $\gamma_0 \in \mathcal{U}$, in such a way the following diagram arise

$$\begin{tikzcd} \mathbb{PN}(C) \arrow[r, "\sigma"] & \mathcal{U} \\ \mathbb{N}^+(C) \arrow[u, "\pi"] \arrow[ur, "\sigma \circ \pi"'] & \end{tikzcd} \tag{3.4}$$

Also consider the canonical projection $\pi_M^{\mathbb{N}^+} : \mathbb{N}^+ \to M$ and the exponential map $\exp : (-\delta,\delta) \times \mathbb{N}^+ \to M$ defined by $\exp(t,v) = \exp_{\pi_M^{\mathbb{N}^+}(v)}(tv)$. Fix $\epsilon > 0$ such that $\Gamma(s) \in \mathcal{U}$ for all $s \in (-\epsilon,\epsilon)$ and let $z : \mathbb{PN}(C) \to \mathbb{N}^+(C)$ be a section of π that, without restriction of generality, can be considered a global section due to the locality of π. Naming $v(s) = z \circ \sigma^{-1} \circ \Gamma(s)$ for $s \in (-\epsilon,\epsilon)$, then we can define a variation $\mathbf{x} : (-\epsilon,\epsilon) \times (-\delta,\delta) \to M$ by $\mathbf{x}(s,t) = \exp(t, v(s)) = \exp_{\pi_M^{\mathbb{N}^+}(v(s))}(tv(s))$. By construction as a composition of differentiable maps, $\mathbf{x}$ is differentiable. Moreover, since $v(s)$ is the initial vector of the geodesic $\gamma_s^{\mathbf{x}}$ defined by $\mathbf{x}(s,t) = \gamma_s^{\mathbf{x}}(t)$, then

$$\gamma_s^{\mathbf{x}} = \sigma \circ \pi(v(s)) = \sigma \circ \pi \circ z \circ \sigma^{-1} \circ \Gamma(s) = \sigma \circ \sigma^{-1} \circ \Gamma(s) = \Gamma(s)$$

for all $s \in (-\epsilon, \epsilon)$, and the lemma follows. □

Lemma 3.6 *Given a variation* $\mathbf{x} : (-\epsilon, \epsilon) \times (-\delta, \delta) \to M$ *by light rays such that* $\mathbf{x}(s,t) = \gamma_s^{\mathbf{x}}(t)$*, then the curve* $\Gamma^{\mathbf{x}} : I \to \mathcal{N}$ *verifying* $\Gamma^{\mathbf{x}}(s) = \gamma_s^{\mathbf{x}}$ *is differentiable.*

Proof. Let $\mathbf{x} : (-\epsilon, \epsilon) \times (-\delta, \delta) \to M$ be a variation by light rays such that $\gamma_s^{\mathbf{x}}(t) = \mathbf{x}(s,t)$. Then the curve

$$\lambda(s) = d\mathbf{x}_{(s,0)} \left(\frac{\partial}{\partial t} \right)_{(s,0)} \in \mathbb{N}^+$$

is clearly differentiable. If $p_{\mathbb{N}^+} : \mathbb{N}^+ \to \mathcal{N}$ is the submersion of proposition 2.2, then $p_{\mathbb{N}^+} \circ \lambda : I \to \mathcal{N}$ is differentiable in $\mathcal{N}$, by composition of differentiable maps. Since

$$p_{\mathbb{N}^+} \circ \lambda(s) = p_{\mathbb{N}^+}\left((\gamma_s^{\mathbf{x}})'(0) \right) = \gamma_s^{\mathbf{x}} = \Gamma^{\mathbf{x}}(s).$$

then $\Gamma^{\mathbf{x}}$ is also differentiable. □

Let us adopt the notation used in lemma 3.6 and call $\Gamma^{\mathbf{x}}$ the curve in $\mathcal{N}$ defined by the variation $\mathbf{x}$ by light rays such that if $\mathbf{x}(s,t) = \gamma_s^{\mathbf{x}}(t)$ then $\Gamma^{\mathbf{x}}(s) = \gamma_s^{\mathbf{x}} \in \mathcal{N}$.

Although the variations defined in lemma 3.5 are not unique, lemma 3.7 shows that all they define the same Jacobi field except by a term in the direction of γ'.

Lemma 3.7 *Let* $\overline{\mathbf{x}} : I \times \overline{H} \to M$ *and* $\mathbf{x} : I \times H \to M$ *be variations by light rays such that* $\Gamma^{\overline{\mathbf{x}}}(s) = \gamma_s^{\overline{\mathbf{x}}}$ *and* $\Gamma^{\mathbf{x}}(s) = \gamma_s^{\mathbf{x}}$ *with* $\gamma_0^{\overline{\mathbf{x}}} = \gamma_0^{\mathbf{x}} = \gamma \in \mathcal{N}$*. Let us denote by* $\overline{J}$ *and* J *the Jacobi fields over* γ *of* $\overline{\mathbf{x}}$ *and* $\mathbf{x}$ *respectively. If* $\Gamma^{\overline{\mathbf{x}}} = \Gamma^{\mathbf{x}}$ *then* $\overline{J} = J\,(\mathrm{mod}\gamma')$.

Proof. We have that $\overline{\mathbf{x}}(s,t) = \gamma_s^{\overline{\mathbf{x}}}(t)$ and $\mathbf{x}(s,\tau) = \gamma_s^{\mathbf{x}}(\tau)$. By lemma 3.4, we can assume without any lack of generality, that $\gamma_s^{\overline{\mathbf{x}}}$ are null geodesics for the metric $\mathbf{g} \in \mathcal{C}$ giving new parameters if necessary. If $\Gamma^{\overline{\mathbf{x}}} = \Gamma^{\mathbf{x}}$ then $\gamma_s^{\overline{\mathbf{x}}} = \gamma_s^{\mathbf{x}}$ for all $s \in I$. Then there exist a differentiable function $h_s(t) = h(s,t)$ such that $\overline{\mathbf{x}}(s,t) = \mathbf{x}(s, h(s,t))$. Hence we have that

$$\frac{\partial \overline{\mathbf{x}}(s,t)}{\partial s} = \frac{\partial \mathbf{x}(s,h(s,t))}{\partial s} + \frac{\partial h(s,t)}{\partial s} \cdot \frac{\partial \mathbf{x}(s,h(s,t))}{\partial \tau}$$

then if $s = 0$

$$\overline{J}(t) = J(h(0,t)) + \frac{\partial h}{\partial s}(0,t) \cdot \gamma'(t)$$

therefore $\overline{J} = J\,(\mathrm{mod}\gamma')$. □

We can wonder how a Jacobi field changes when another metric of the same conformal class is considered in M. The following result shows it with a proof similar to the one of lemma 3.7.

Lemma 3.8 *Let $\mathbf{x} : I \times H \to M$ be a variation by light rays of $\gamma = \mathbf{x}(0,\cdot)$. If $J \in \mathcal{J}_L(\gamma)$ is the Jacobi field of $\mathbf{x}$ along γ related to the metric $\mathbf{g} \in \mathcal{C}$, then the Jacobi field $\overline{J}$ of $\mathbf{x}$ along γ related to another metric $\overline{\mathbf{g}} \in \mathcal{C}$ verifies*

$$\overline{J} = J\,(\mathrm{mod}\gamma')\,.$$

Proof. Let $\mathbf{x} : I \times H \to M$ be a variation by light rays of γ where $\mathbf{x}(s,t) = \gamma_s(t)$ with $\gamma = \gamma_0$. By lemma 3.4, we can assume that γ_s is null geodesic related to the metric $\mathbf{g}$ and there exists changes of parameters $h_s : \overline{H} \to H$ such that $\overline{\gamma_s}(\tau) = \gamma_s(h_s(\tau))$ are null geodesics related to $\overline{\mathbf{g}} \in \mathcal{C}$ for all $s \in I$ and where $h(s,\tau) = h_s(\tau)$ is a differentiable function. So, we consider that $J \in \mathcal{J}_L(\gamma)$ is the Jacobi field of $\mathbf{x}$ along γ and $\overline{J} \in \overline{\mathcal{J}}_L(\gamma)$ is the one of $\overline{\mathbf{x}}$. Then $\overline{\mathbf{x}}(s,\tau) = \mathbf{x}(s,h_s(\tau))$ and we have that

$$\begin{aligned}\overline{J}(\tau) &= \left.\frac{\partial \overline{\mathbf{x}}(s,\tau)}{\partial s}\right|_{(0,\tau)} = \left.\frac{\partial \mathbf{x}(s,h_s(\tau))}{\partial s}\right|_{(0,\tau)} = \\ &= \frac{\partial \mathbf{x}}{\partial s}(0,h_0(\tau)) + \frac{\partial h}{\partial s}(0,\tau)\frac{\partial \mathbf{x}}{\partial t}(0,h_0(\tau)) = J(h_0(\tau)) + \frac{\partial h}{\partial s}(0,\tau)\,\gamma'(h_0(\tau))\end{aligned}$$

therefore $\overline{J} = J\,(\mathrm{mod}\gamma')$. □

Lemma 3.9 *Given two variations by light rays $\mathbf{x} : I \times H \to M$ and $\overline{\mathbf{x}} : \overline{I} \times \overline{H} \to M$ such that $\Gamma^{\mathbf{x}}(0) = \Gamma^{\overline{\mathbf{x}}}(0) = \gamma$. Let us denote by J and $\overline{J}$ their corresponding Jacobi fields at $0 \in I$ and $0 \in \overline{I}$ of $\mathbf{x}$ and $\overline{\mathbf{x}}$ respectively. If $(\Gamma^{\mathbf{x}})'(0) = \left(\Gamma^{\overline{\mathbf{x}}}\right)'(0)$ then $J = \overline{J}\,(\mathrm{mod}\gamma')$.*

Proof. Due to we want to compare the Jacobi fields J and $\overline{J}$ on γ, we can assume without any lack of generality that $\mathbf{x}$ as well as $\overline{\mathbf{x}}$ provide the same geodesic parameter for γ, then by lemmas 3.5 and 3.7, we can consider that $\mathbf{x}(s,t) = \exp_{\alpha(s)}(tu(s))$ and $\overline{\mathbf{x}}(r,t) = \exp_{\overline{\alpha}(r)}(t\overline{u}(r))$ where $u = u(0) = \overline{u}(0)$ and also $p = \alpha(0) = \overline{\alpha}(0)$.

Moreover, we can assume the diagram 3.4 holds.

Since $(\Gamma^{\mathbf{x}})'(0) = \left(\Gamma^{\overline{\mathbf{x}}}\right)'(0)$ then we have

$$d\sigma_{[v(0)]} \circ d\pi_{v(0)}(v'(0)) = d\sigma_{[\overline{v}(0)]} \circ d\pi_{\overline{v}(0)}(\overline{v}'(0)) \Leftrightarrow d\pi_{v(0)}(v'(0)) = d\pi_{\overline{v}(0)}(\overline{v}'(0))$$

Observe that $[v(0)] = [\overline{v}(0)]$ and thus, $d\pi_{v(0)} = d\pi_{\overline{v}(0)}$, and its kernel is the subspace generated by the tangent vector at $s = 0$ of the curve $c(s) = e^s v(0)$, hence

$$v'(0) = \overline{v}'(0) + \mu c'(0) \tag{3.5}$$

with $\mu \in \mathbb{R}$. By remark 3.2, we have that

$$\begin{cases} \alpha'(0) = \overline{\alpha}'(0) \\ \frac{Du}{ds}(0) = \frac{D\overline{u}}{dr}(0) + \mu\frac{Dc}{ds}(0) \end{cases} \Rightarrow \begin{cases} \alpha'(0) = \overline{\alpha}'(0) \\ \frac{Du}{ds}(0) = \frac{D\overline{u}}{dr}(0) + \mu\gamma'(0) \end{cases}$$

therefore we conclude that $J = \overline{J}\,(\mathrm{mod}\gamma')$. □

The differentiable structure of $\mathcal{N}$ has been built in section 2 from the one in $\mathbb{PN}(C)$ where C is a local spacelike Cauchy surface. So, we will identify the tangent space $T_\gamma\mathcal{N}$ with some quotient space of $\mathcal{J}_L(\gamma)$ via a tangent space of $\mathbb{PN}(C)$.

Proposition 3.4 *Given $\xi \in T_{\gamma_{u_0}}\mathcal{N}$ such that $\Gamma'(0) = \xi$ for some curve $\Gamma \subset \mathcal{N}$. Let $\mathbf{x} = \mathbf{x}(s,t)$ be a variation by light rays of γ_{u_0} verifying that $\Gamma^{\mathbf{x}} = \Gamma$ such that $J \in \mathcal{L}(\gamma_{u_0})$ is the Jacobi field over γ_{u_0} of $\mathbf{x}$. If $\zeta : T_{\gamma_{u_0}}\mathcal{N} \to \mathcal{L}(\gamma_{u_0})$ is the map defined by*

$$\overline{\zeta}(\xi) = J\left(\mathrm{mod}\gamma'_{u_0}\right)$$

then $\overline{\zeta}$ is well–defined and a linear isomorphism.

Proof. By lemma 3.9, $\overline{\zeta}$ is well–defined.

We have seen in section 2 that for a globally hyperbolic open set $V \subset M$ such that $C \subset V$ is a smooth local spacelike Cauchy surface, the diagram 2.3 given by

$$\mathcal{N} \supset \mathcal{U} \simeq \mathbb{PN}(C) \simeq \Omega^X(C) \hookrightarrow \mathbb{N}^+(C) \hookrightarrow \mathbb{N}^+ \hookrightarrow TM$$

holds. Proposition 3.3 shows that $\zeta : T_uTM \to \mathcal{J}(\gamma_u)$ is a linear isomorphism for any $u \in TM$.

The idea of the proof is the following. We will restrict ζ from T_uTM up to $T_{[u]}\mathbb{PN}(C)$ step by step, identifying the corresponding subspace of $\mathcal{J}(\gamma_u)$ image of the map. In the first step we obtain that $T_{u_0}\mathbb{N}^+ \to \mathcal{J}_L(\gamma_{u_0})$ is a isomorphism. In the next step we obtain the isomorphism $T_{u_0}\mathbb{N}^+(C) \to S$ where $S \subset \mathcal{J}_L(\gamma_{u_0})$ is of the same codimension and transverse (that is, linearly independent) to the vector subspace $\widehat{\mathcal{J}}'_0(\gamma_{u_0})$, then S is isomorphic to $\mathcal{J}_L(\gamma_{u_0})/\widehat{\mathcal{J}}'_0(\gamma_{u_0})$. Next, we consider the isomorphism $T_{[u_0]}\mathbb{PN}(C) \to S/\left(S \cap \widehat{\mathcal{J}}_0(\gamma_{u_0})\right)$, but since $\mathcal{U} \simeq \mathbb{PN}(C)$, then

$$T_{\gamma_{u_0}}\mathcal{N} \to S/\left(S \cap \widehat{\mathcal{J}}_0(\gamma_{u_0})\right)$$

is an isomorphism.

Recall that we have denoted $\mathcal{J}_0(\gamma_{u_0}) = \widehat{\mathcal{J}}_0(\gamma_{u_0}) \oplus \widehat{\mathcal{J}}'_0(\gamma_{u_0})$. Observe that the linear map $q : S \to \mathcal{J}_L(\gamma_{u_0})/\mathcal{J}_0(\gamma_{u_0})$ defined by $q(J) = [J]$ verifies that

$$q(J) = [0] \Leftrightarrow J(t) = (a+bt)\,\gamma'_{u_0}(t) \Leftrightarrow J \in S \cap \widehat{\mathcal{J}}_0(\gamma_{u_0})$$

then $S/\left(S \cap \widehat{\mathcal{J}}_0(\gamma_{u_0})\right)$ is isomorphic to $\mathcal{L}(\gamma_{u_0}) = \mathcal{J}_L(\gamma_{u_0})/\mathcal{J}_0(\gamma_{u_0})$. This shows that the map $\overline{\zeta} : T_{\gamma_{u_0}}\mathcal{N} \to \mathcal{L}(\gamma_{u_0}) = \mathcal{J}_L(\gamma_{u_0})/\mathcal{J}_0(\gamma_{u_0})$, $\xi \mapsto [J]$, is a linear isomorphism and the proof is complete. □

Proposition 3.4 allows to see the vectors of the tangent space $T_\gamma\mathcal{N}$ as Jacobi fields of variations by light rays. We will use, from now on, this characterization when working with tangent vectors of $\mathcal{N}$.

By propositions 2.1 and proposition 3.4, it is clear that the characterization of $T_\gamma \mathcal{N}$ as $\mathcal{L}(\gamma)$ does not depends on the representative of the conformal class $\mathcal{C}$.

4. The canonical contact structure in $\mathcal{N}$

In this section, the canonical contact structure on $\mathcal{N}$ will be discussed. Such contact struture is inherited from the kernel of the *canonical 1–form* of T^*M and it will be described by passing the distribution of hyperplanes to TM before pushing it down to $\mathcal{N}$ in virtue of the inclusions of eq. 2.3. The basic elements of symplectic and contact geometry can be consulted in [Ab87], [Ar89] and [LM87].

4.1. Elements of symplectic geometry in T^*M

Consider a differentiable manifold M. Its cotangent bundle $\pi\colon T^*M \to M$ carries a canonical 1–form θ defined pointwise at every $\alpha \in T^*M$ by $\theta_\alpha = (d\pi_\alpha)^* \alpha$. Consequently we have

$$\theta_\alpha(\xi) = \left((d\pi_\alpha)^* \alpha\right)(\xi) = \alpha\left((d\pi_\alpha)\xi\right) \tag{4.1}$$

for $\xi \in T_\alpha(T^*M)$. In local canonical bundle coordinates (x^k, p_k), we can write

$$\theta = \sum_{k=1}^{m} p_k dx^k . \tag{4.2}$$

The 2–form $\omega = -d\theta$, defines a symplectic 2–form in T^*M, that in the previous local coordinates takes the form $\omega = \sum_{k=1}^{m} dx^k \wedge dp_k$.

Now, we want to construct $\mathcal{N}$ again, but this time starting from the contangent bundle T^*M. Consider again the natural identification provided by the metric $\mathbf{g}$: $\widehat{\mathbf{g}} : TM \to T^*M$, $\xi \mapsto \mathbf{g}(\xi, \cdot)$, and denote by $\mathbb{N}^{+*}$ the image of the restriction of $\widehat{\mathbf{g}}$ to $\mathbb{N}^+$, that is

$$\mathbb{N}^{+*} = \widehat{\mathbf{g}}\left(\mathbb{N}^+\right) = \left\{\alpha = \widehat{\mathbf{g}}(\xi) \in T^*M : \xi \in \mathbb{N}^+\right\}$$

In an analogous manner as in Sect. 2, define the *Euler field* $\mathcal{E} \in \mathfrak{X}(T^*M)$ by

$$\mathcal{E}(\alpha) = dc\left(\frac{\partial}{\partial t}\right)(0) ,$$

where $\alpha \in T_p^*M$ and $c : \mathbb{R} \to T_p^*M$ verifies that $c(t) = e^t \alpha$. The curve c is an integral curve of $\mathcal{E}$ because

$$c'(t) = dc\left(\frac{\partial}{\partial t}\right)(t) = \mathcal{E}(c(t)) .$$

In the previous coordinates, $\mathcal{E}$ can be written as $\mathcal{E} = p_k \partial/\partial p_k$. So, for every $\alpha \in \mathbb{N}^{+*}$ the integral curve $c(t) = e^t \alpha$ is contained in $\mathbb{N}^{+*}$, therefore $\mathcal{E}$ is tangent to $\mathbb{N}^{+*}$.

Moreover, if ω is the symplectic 2–form of T^*M it is trivial to see

$$\mathcal{L}_{\mathcal{E}}\omega = i_{\mathcal{E}} d\omega + d(i_{\mathcal{E}}\omega) = d(-\theta) = -d\theta = \omega , \tag{4.3}$$

therefore $\mathcal{E}$ is a Liouville vector field. In fact, $\mathcal{E}$ sometimes is called the *Liouville* or *Euler–Liouville vector field*.

Consider now the Hamiltonian function (again just the kinetic energy) defined by $H : T^*M \to \mathbb{R}$, $\alpha \mapsto \frac{1}{2}\mathbf{g}\left(\widehat{\mathbf{g}}^{-1}(\alpha), \widehat{\mathbf{g}}^{-1}(\alpha)\right)$, defining the Hamiltonian vector field:

$$X_H = g^{ki} p_i \frac{\partial}{\partial x^k} - \frac{1}{2}\frac{\partial g^{ij}}{\partial x^k} p_i p_j \frac{\partial}{\partial p_i}$$

Lemma 4.1 *Let $X_{\mathbf{g}}, \Delta \in \mathfrak{X}(TM)$ be the the geodesic spray and Euler field of TM and $X_H, \mathcal{E} \in \mathfrak{X}(T^*M)$ the Hamiltonian vector field and Euler field of T^*M respectively. Then we have that $\widehat{\mathbf{g}}_*(\Delta) = \mathcal{E}$ and $\widehat{\mathbf{g}}_*(X_{\mathbf{g}}) = X_H$.*

Proof. If we take any $\xi \in T^*M$ and $\alpha = \widehat{\mathbf{g}}(\xi)$, then the integral curve $c(t) = e^t\xi$ of Euler field Δ in TM is transformed by $\widehat{\mathbf{g}}$ as

$$\widehat{\mathbf{g}}(c(t)) = \mathbf{g}(c(t), \cdot) = \mathbf{g}\left(e^t\xi, \cdot\right) = e^t\mathbf{g}(\xi, \cdot) = e^t\widehat{\mathbf{g}}(\xi) = e^t\alpha \in T^*M$$

being an integral curve of Euler field $\mathcal{E}$ in T^*M. Then, for any $\xi \in T^*M$ we have that

$$\widehat{\mathbf{g}}_*(\Delta(\xi)) = \mathcal{E}(\widehat{\mathbf{g}}(\xi))$$

is verified, therefore this implies $\widehat{\mathbf{g}}_*(\Delta) = \mathcal{E}$.

The second relation is obtained easily by taking the pull-back of the identity $i_{X_H}\omega = dH$ along the map $\hat{\mathbf{g}}$. $\square$

The following corollary is an immediate consequence of Lemma 4.1 and the construction of $\mathcal{N}$ done in section 2.

Corollary 4.1 *The space of light rays $\mathcal{N}$ of M can be built by the quotient*

$$\mathcal{N} = \mathbb{N}^{+*}/\mathcal{D}^*$$

where $\mathcal{D}^*$ is the distribution generated by the vector fields $\mathcal{E}$ and X_H, that is $\mathcal{D}^* =$ $\operatorname{span}\{\mathcal{E}, X_H\}$.

Lemma 4.1 also shows that the null geodesic defined by $\alpha \in \mathbb{N}^*$ coincides to the null geodesic defined by $v \in \mathbb{N}$ if and only if $\widehat{\mathbf{g}}(v) = \alpha$, because the first equation has to be verified. Then we have the following commutative diagram:

$$\begin{tikzcd} \mathbb{N}^* \arrow[r, "p_{\mathbb{N}^*}"] & \mathcal{N} \\ \mathbb{N} \arrow[u, "\widehat{\mathbf{g}}"] \arrow[ur, "p_{\mathbb{N}}"'] & \end{tikzcd} \tag{4.4}$$

Next, we will introduce some basic definitions and results in contact geometry that we will need later. See [Ar89, Appx. 4] and [LM87, Ch. 5] for more details.

Definition 4.1 *Given a n–dimensional differentiable manifold P, a* contact element *in P is a $(n-1)$–dimensional subspace $\mathcal{H}_q \subset T_qP$. The point $q \in P$ is called the* contact point *of $\mathcal{H}_q$.*

We will say that a distribution of hyperplanes *$\mathcal{H}$ in a differentiable manifold M is a map $\mathcal{H}$ defined in M such that for every $q \in M$ we have that $\mathcal{H}(q) = \mathcal{H}_q$ is a contact element at q.*

Lemma 4.2 *Every differentiable distribution of hyperplanes $\mathcal{H}$ can be written locally as the kernel of 1–form.*

Proof. See [G08, Lem. 1.1.1] for proof. □

It is clear that if a differentiable distribution of hyperplanes $\mathcal{H}$ is defined locally by the 1–form $\alpha \in \mathfrak{X}^*(P)$ then, for every non-vanishing function $f \in \mathfrak{F}(P)$ the 1–form $f\alpha$ also defines $\mathcal{H}$ since α and $f\alpha$ have the same kernel.

Given a distribution of hyperplanes $\mathcal{H}$ we will say that it is maximally non-integrable if for any locally defined 1-form η such that $\mathcal{H} = \ker\eta$, then $d\eta$ is non-degenerate when restricted to $\mathcal{H}$.

Definition 4.2 *A* contact structure *$\mathcal{H}$ in a $(2n+1)$–dimensional differentiable manifold P is a maximally non–integrable distribution of hyperplanes. The hyperplanes $\mathcal{H}_x \subset T_xP$ are called contact elements. If there exists a globally defined 1-form η defining $\mathcal{H}$, i.e., $\mathcal{H} = \ker\eta$, we will say that $\mathcal{H}$ is a* cooriented contact structure *and we will say that η is a contact form.*

An equivalent way to determine if a distribution of hyperplanes $\mathcal{H}$ determines a contact structure is provided by the following result (see also [Ar89] and [Ca01]).

Lemma 4.3 *Let $\mathcal{H}$ be a distribution of hyperplanes in P locally defined as $\mathcal{H} = \ker(\eta)$, then $d\eta|_{\mathcal{H}}$ is non–degenerated if and only if $\eta \wedge (d\eta)^n \neq 0$.*

Proof. See [Ca01, Prop. 10.3] for proof. □

Lemma 4.4 *If α is a contact form in P, then $f\alpha$ is also a contact form for every non–vanishing differentiable function $f \in \mathfrak{F}(P)$.*

Proof. See [LM87, Sect. V.4.1]. □

4.2. Constructing the contact structure of $\mathcal{N}$

Consider the tautological 1–form $\theta \in \mathfrak{X}^*(T^*M)$. The diffeomorphism $\widehat{\mathbf{g}} : TM \to T^*M$ allows to carry away θ to TM by pull–back. Let $\pi_M^{TM} : TM \to M$ and $\pi_M^{T^*M} : T^*M \to M$ be the canonical projections, since $\pi_M^{TM} = \pi_M^{T^*M} \circ \widehat{\mathbf{g}}$, then it is verified

$$\left(d\pi_M^{TM}\right)_v(\xi) = \left(d\pi_M^{T^*M}\right)_{\widehat{\mathbf{g}}(v)}\left(\widehat{\mathbf{g}}_*(\xi)\right)$$

for all $\xi \in T_v TM$. If we define

$$\theta_{\mathbf{g}} = \widehat{\mathbf{g}}^* \theta \tag{4.5}$$

then, using the expression 4.1, if $\xi \in T_v TM$ we have

$$(\theta_{\mathbf{g}})_v (\xi) = \widehat{\mathbf{g}}(v) \left(\left(d\pi_M^{T^*M} \right)_{\widehat{\mathbf{g}}(v)} (\widehat{\mathbf{g}}_* (\xi)) \right) = \mathbf{g} \left(v, \left(d\pi_M^{TM} \right)_v (\xi) \right) .$$

For a given globally hyperbolic open set $V \subset M$ equipped with coordinates $\left(x^1, \ldots, x^m\right)$ such that $v \in TV$ is written as $v = v^i \frac{\partial}{\partial x^i}$, then $\left(x^i, v^i\right)$ are coordinates in TV. By expression 4.2, we can write

$$\theta_{\mathbf{g}} = g_{ij} v^i dx^j .$$

Let us denote by $\mathcal{H}^{TV} = \ker(\theta_{\mathbf{g}})$, that is a distribution of hyperplanes in $TV \subset TM$. This implies that $\dim\left(\mathcal{H}_v^{TV}\right) = 2m-1$ for every $v \in TV$.

As we seen in Sect. 2, we have the chain of inclusions 2.3:

$$\Omega \hookrightarrow \mathbb{N}^+(C) \hookrightarrow \mathbb{N}^+(V) \hookrightarrow TV \tag{4.6}$$

where $\Omega = \Omega^T(C) = \{v \in \mathbb{N}^+ \mid g(v,T) = -1\}$ for a non-vanishing timelike vector field T. Observer taht if $v \in \Omega$ is the representative of the class of equivalence $[v] \in \mathbb{PN}(C)$, then clearly the following maps

$$\begin{array}{ccccc} \Omega & \longrightarrow & \mathbb{PN}(C) & \longrightarrow & \mathcal{U} \subset \mathcal{N} \\ v & \mapsto & [v] & \mapsto & \gamma_v \end{array} \tag{4.7}$$

are diffeomorphisms.

Then, we will see that the pullback of $\theta_{\mathbf{g}}$ by the inclusion $\Omega \hookrightarrow TV$ defines a 1–form $\theta_{\mathbf{g}}|_{\Omega^X(C)}$, and therefore a distribution of hyperplanes, in Ω. This 1–form and its kernel can be extended from $\mathcal{U} \subset \mathcal{N}$ obtaining the 1–form θ_0 looked for.

To obtain a suitable formula of θ_0 we will proceed projecting the distribution of hyperplanes in TM up to $\Omega^X(C)$ step by step.

First, observe that the restriction of $\mathcal{H}^{TV}$ to $T\mathbb{N}^+(V)$, denoted by $\mathcal{H}^{\mathbb{N}^+(V)}$, is again a distribution of hyperplanes. Indeed, if $c : (-\epsilon, \epsilon) \to \mathbb{N}^+(V)$ is a differentiable curve such that

$$\left\{ \begin{array}{l} \alpha(s) = \pi_M^{\mathbb{N}^+}(c(s)) \text{ is a timelike curve} \\ v = c(0) \in \mathbb{N}^+(V) \\ \xi = c'(0) \in T_v \mathbb{N}^+(V) \end{array} \right.$$

then

$$\theta_{\mathbf{g}}(\xi) = \mathbf{g}(v, \alpha'(0)) \neq 0$$

since v is null and $\alpha'(0)$ timelike. This implies that $\xi \notin \mathcal{H}_v^{TV}$. So, we have that $T_v TV = span\{\xi\} \oplus \mathcal{H}_v^{TV}$ and since $span\{\xi\} \subset T_v\mathbb{N}^+(V)$ and $\mathcal{H}_v^{\mathbb{N}^+(V)} = \mathcal{H}_v^{TV} \cap T_v\mathbb{N}^+(V)$ then we have that

$$\dim\left(\mathcal{H}_v^{\mathbb{N}^+(V)}\right) = 2m-2$$

therefore $\mathcal{H}^{\mathbb{N}^+(V)}$ is a distribution of hyperplanes in $\mathbb{N}^+(V)$.

The next step is to restrict $\mathcal{H}^{\mathbb{N}^+(V)}$ to $T\mathbb{N}^+(C)$, where C is a Cauchy surface of V. Again, as done above, if $\gamma : I \to M$ is a null geodesic verifying $\gamma(0) \in C$ and $\gamma'(0) = v \in \mathbb{N}^+(C)$, since the vector subspace $Z = \{u \in T_vM : \mathbf{g}(v,u) = 0\}$ is $m-1$–dimensional and $v = \gamma'(0) \in Z$, then $\dim\left(Z \cap T_{\gamma(0)}C\right) = m-2$. Hence, we can pick up a vector $\eta \in T_{\gamma(0)}C$ such that $T_{\gamma(0)}C = span\{\eta\} \oplus \left(Z \cap T_{\gamma(0)}C\right)$. Now, we can choose a differentiable curve $c : (-\epsilon,\epsilon) \to \mathbb{N}^+(C)$ verifying

$$\begin{cases} c(0) = v \in \mathbb{N}^+(C) \\ c'(0) = \kappa \in T_v\mathbb{N}^+(C) \\ \left(d\pi_M^{\mathbb{N}^+}\right)_v(\kappa) = \lambda\eta \text{ for } \lambda \neq 0 \end{cases}$$

then

$$\theta_{\mathbf{g}}(\kappa) = \mathbf{g}\left(v, \left(d\pi_M^{\mathbb{N}^+}\right)_v(\kappa)\right) = \mathbf{g}(v,\lambda\eta) \neq 0$$

because $\eta \notin Z$, and this shows that $\kappa \notin \mathcal{H}_v^{\mathbb{N}^+(V)}$. Then $T_v\mathbb{N}^+(V) = span\{\kappa\} \oplus \mathcal{H}_v^{\mathbb{N}^+(V)}$ and since $span\{\kappa\} \subset T_v\mathbb{N}^+(C)$ and $\mathcal{H}_v^{\mathbb{N}^+(C)} = \mathcal{H}_v^{\mathbb{N}^+(V)} \cap T_v\mathbb{N}^+(C)$, then it follows

$$\dim\left(\mathcal{H}_v^{\mathbb{N}^+(C)}\right) = \dim\left(T_v\mathbb{N}^+(C)\right) - 1 = 2m-3$$

thus $\mathcal{H}^{\mathbb{N}^+(C)}$ is a distribution of hyperplanes in $\mathbb{N}^+(C)$.

It is possible to repeat the previous argument to show that the restriction of $\mathcal{H}^{\mathbb{N}^+(C)}$ to $T\Omega$ defines a distribution of hyperplanes. In fact, consider some $\eta \in T_{\gamma(0)}C$ in the same condition as before and take a differentiable curve $c : (-\epsilon,\epsilon) \to \Omega$ verifying

$$\begin{cases} c(0) = v \in \Omega \\ c'(0) = \kappa \in T_v\Omega \\ \left(d\pi_M^{\mathbb{N}^+}\right)_v(\kappa) = \lambda\eta \text{ for } \lambda \neq 0 \end{cases}$$

then again

$$\theta_{\mathbf{g}}(\kappa) = \mathbf{g}(v,\lambda\eta) \neq 0$$

showing that $\kappa \notin \mathcal{H}_v^{\mathbb{N}^+(C)}$. Then $T_v\mathbb{N}^+(C) = \mathrm{span}\{\kappa\} \oplus \mathcal{H}_v^{\mathbb{N}^+(C)}$ and since $\mathrm{span}\{\kappa\} \subset T_v\Omega$ then we have that

$$\dim\left(\mathcal{H}_v^{\Omega}\right) = \dim\left(T_v\Omega\right) - 1 = 2m-4$$

thus $\mathcal{H}^{\Omega}$ is a distribution of hyperplanes in $\Omega \subset \mathbb{N}^+(C)$.

By this process of restriction from TV to Ω we have passed $\mathcal{H}^{TV} \subset TTV$ as a distribution of hyperplanes $\mathcal{H}^{\Omega} \subset T\Omega \subset TTV$. Moreover since $\mathcal{H}^{TV} = \ker(\theta_{\mathbf{g}})$ and $\mathcal{H}^{\Omega} = T\Omega \cap \mathcal{H}^{TV}$ then

$$\mathcal{H}^{\Omega} = \ker\left(\left.\theta_{\mathbf{g}}\right|_{\Omega}\right)$$

where $\theta_{\mathbf{g}}|_{\Omega}$ denotes the restriction of $\theta_{\mathbf{g}}$ to Ω. This fact is important in order to show that $\mathcal{H}^{\Omega}$ is a contact structure.

Then, using the diffeomorphisms in (4.7), $\mathcal{H}^{\Omega}$ passes to $\mathcal{U} \subset \mathcal{N}$ as a distribution of hyperplanes of dimension $2m-4$. Let us denote by $\mathcal{H} \subset T\mathcal{N}$ said distribution.

Proposition 4.1 *If $\mathcal{U} \subset \mathcal{N}$ and $T \in \mathfrak{X}(M)$ is a given global non-vanishing timelike vector field as above, then the distribution of hyperplanes*

$$\mathcal{H}(\mathcal{U}) = \{[J] \in T_{\gamma}\mathcal{U} : \mathbf{g}(\gamma'(0), J(0)) = 0 \text{ with } \mathbf{g}(\gamma'(0), T) = -1\} \tag{4.8}$$

is a contact structure.

Proof. Since $\omega = -d\theta$, then taking the exterior derivative on $\theta_{\mathbf{g}}$ we obtain

$$\omega_{\mathbf{g}} = -d\theta_{\mathbf{g}}, \tag{4.9}$$

therefore we have

$$\omega_{\mathbf{g}} = g_{ij} dx^j \wedge dv^i + \frac{\partial g_{ij}}{\partial x^k} v^i dx^j \wedge dx^k \tag{4.10}$$

that clearly shows that $\omega_{\mathbf{g}}$ is a symplectic 2–form in TM (notice that $\omega_{\mathbf{g}}^n = \det(g_{ij})\, dx^1 \wedge \cdots \wedge dx^n \wedge dv^1 \wedge \cdots \wedge dv^n \neq 0$).

Consider two curves $u_n(s) = u_n^i(s) \left(\frac{\partial}{\partial x^i}\right)_{\alpha_n(s)} \in TM$ where $n = 1, 2$ such that

$$\begin{aligned} \alpha_n'(s) &= a_n^i(s) \left(\tfrac{\partial}{\partial x^i}\right)_{\alpha_n(s)} \\ u_n'(s) &= a_n^i(s) \left(\tfrac{\partial}{\partial x^i}\right)_{u_n(s)} + \tfrac{du_n^i}{ds}(s) \left(\tfrac{\partial}{\partial v^i}\right)_{u_n(s)} \end{aligned}$$

and recall that

$$\frac{Du_n}{ds} = \left(\frac{du_n^k}{ds} + \Gamma_{ij}^k a_n^i u_n^j\right) \left(\frac{\partial}{\partial x^k}\right)_{\alpha_n}$$

calling $\frac{D^k u_n}{ds} = \frac{du_n^k}{ds} + \Gamma_{ij}^k a_n^i u_n^j$ the k–th component of $\frac{Du_n}{ds}$. If $u = u_1(0) = u_2(0)$ and $\xi_n = u_n'(0)$ for $n = 1, 2$, then we have that:

$$\begin{aligned} \omega_{\mathbf{g}}(\xi_1, \xi_2) &= g_{ij} a_1^i \frac{D^j u_2}{ds} - g_{ij} a_2^j \frac{D^i u_1}{ds} + \left(g_{kl}\Gamma_{ji}^l - g_{jl}\Gamma_{ki}^l + \frac{\partial g_{ij}}{\partial x^k} - \frac{\partial g_{ik}}{\partial x^j}\right) u^i a_1^j a_2^k = \\ &= g_{ij} a_1^i \frac{D^j u_2}{ds} - g_{ij} a_2^j \frac{D^i u_1}{ds} = \mathbf{g}\left(\alpha_1'(0), \frac{Du_2}{ds}(0)\right) - \mathbf{g}\left(\alpha_2'(0), \frac{Du_1}{ds}(0)\right). \end{aligned}$$

Since the exterior derivative commutes with the restriction to submanifolds, then

$$\omega_{\mathbf{g}}|_{\Omega} = -\left(d\theta_{\mathbf{g}}\right)|_{\Omega} = -d\left(\theta_{\mathbf{g}}|_{\Omega}\right)$$

Proposition 3.3 permit to transmit $\theta_{\mathbf{g}}|_{\Omega}, \omega_{\mathbf{g}}|_{\Omega}$ to $\mathcal{L}(\gamma_u)$ pointwise. Calling θ_0 and ω_0 the resulting forms, then for $[J], [J_1], [J_2] \in \mathcal{L}(\gamma_u)$ we have

$$\theta_0([J]) = \mathbf{g}(\gamma_u'(0), J(0))$$

where γ_u is parametrized such that $\gamma_u'(0) \in \Omega$, and

$$\omega_0([J_1],[J_2]) = \mathbf{g}(J_1(0), J_2'(0)) - \mathbf{g}(J_2(0), J_1'(0)) \tag{4.11}$$

In order to prove that $\mathcal{H}$ is a contact structure, we will show that $\omega_0|_{\mathcal{H}}$ is non–degenerated. Consider $[J_1],[J_2] \in \mathcal{H}$, then the initial values of J_1 and J_2 in expression 4.11 verify

$$\mathbf{g}(J_i(0), \gamma_u'(0)) = 0 \ ; \qquad \mathbf{g}(J_i'(0), \gamma_u'(0)) = 0 \,. \tag{4.12}$$

for $i = 1, 2$, that is $J_i(0), J_i'(0) \in \{\gamma_u'\}^{\perp} = \left\{v \in T_{\gamma_u(0)}M : \mathbf{g}(v, \gamma_u'(0)) = 0\right\}$.

For a given $[J_1] \in \mathcal{H}$, if $\omega_0([J_1],[J_2]) = 0$ for all $[J_2] \in \mathcal{L}(\gamma_u)$, then in particular, also for $[J_2]$ verifying $J_2'(0) = 0$, we have

$$\omega_0([J_1],[J_2]) = 0 \Rightarrow \mathbf{g}(J_2(0), J_1'(0)) = 0$$

Since $J_1'(0) \in \{\gamma_u'\}^{\perp}$, the only vector $J_1'(0)$ such that $\mathbf{g}(J_2(0), J_1'(0)) = 0$ for all $J_2(0) \in \{\gamma_u'\}^{\perp}$ is, by definition of $\{\gamma_u'\}^{\perp}$, the vector $J_1'(0) = 0 \, (\mathrm{mod} \gamma_u')$.

On the other hand, for $[J_2]$ verifying $J_2(0) = 0$ we have

$$\omega_0([J_1],[J_2]) = 0 \Rightarrow \mathbf{g}(J_1(0), J_2'(0)) = 0$$

and again, since $J_1(0) \in \{\gamma_u'\}^{\perp}$ then the only vector $J_1(0)$ such that $\mathbf{g}(J_1(0), J_2'(0)) = 0$ for all $J_2'(0) \in \{\gamma_u'\}^{\perp}$ is $J_1(0) = 0 \, (\mathrm{mod} \gamma_u')$.

Thus, the only $[J_1] \in \mathcal{H}$ such that $\omega_0([J_1],[J_2]) = 0$ for all $[J_2] \in \mathcal{H}$ is $J_1 = 0 \, (\mathrm{mod} \gamma_u')$, therefore $\omega_0|_{\mathcal{H}}$ is non–degenerated. This shows that $\mathcal{H}$ is a contact structure in $\mathcal{N}$. □

Let us take $\gamma \in \mathcal{U} \cap \mathcal{V}$, since in general $\frac{d}{dt}\mathbf{g}(\gamma'(t), T(\gamma(t))) \neq 0$, then there are different parameter for γ in order to write $\mathcal{H}(\mathcal{U})$ and $\mathcal{H}(\mathcal{V})$ as in expression 4.8. If we consider that $\gamma = \gamma(t)$ and $\overline{\gamma} = \overline{\gamma}(\tau)$ are the parametrizations of $\gamma \in \mathcal{U} \cap \mathcal{V}$ such that $\overline{\gamma}(\tau) = \gamma(a\tau + b)$ verifying

$$\mathbf{g}(\gamma'(0), T) = -1 \ ; \qquad \mathbf{g}(\overline{\gamma}'(0), T) = -1 \,.$$

By definition of $\mathcal{J}_L(\overline{\gamma})$, we have that $\mathbf{g}\left(\overline{J}(\tau), \overline{\gamma}'(\tau)\right)$ is constant, therefore

$$\mathbf{g}\left(\overline{J}(0), \overline{\gamma}'(0)\right) = \mathbf{g}\left(\overline{J}(-b/a), \overline{\gamma}'(-b/a)\right) = a\mathbf{g}(J(0), \gamma'(0))$$

as we have seen in remark 3.1, whence since $\overline{\gamma}(-b/a) = \gamma(0)$ we have

$$\mathbf{g}\left(\overline{J}(-b/a), \overline{\gamma}'(-b/a)\right) = 0 \Leftrightarrow \mathbf{g}(J(0), \gamma'(0)) = 0$$

The same argument above is valid to prove that $\mathcal{H}_\gamma$ does not depends on the timelike vector field used to define Ω, because it only affects to the parametrization of γ. This shows that $\mathcal{H}_\gamma$ is well defined and does not depends on the neighbourhood used in its construction. In addition, using Lemma 3.8, can be shown that this contact structure does not depend on the auxiliary metric $\mathbf{g}$ selected in the conformal class $\mathcal{C}$.

At this point, we may consider a covering $\{\mathcal{U}_\delta\}_{\delta\in I} \subset \mathcal{N}$ and, for any $\delta \in I$, consider the local 1–form θ_0^δ defining the contact structure $\mathcal{H}$ as before. If we take a partition of unity $\{\chi_\delta\}_{\delta\in I}$ subordinated to the covering $\{\mathcal{U}_\delta\}_{\delta\in I}$ then we can define a global 1–form by:

$$\theta_0\left([J]\right) = \sum_{\delta\in I} \chi_\delta\left([J]\right)\cdot\theta_0^\delta\left([J]\right)$$

then the contact structure $\mathcal{H}$ is cooriented since θ_0 is globally defined and, by Lemma 4.3, remains maximally non-integrable.

In the following section we will provide a slightly more intrinsic construction of the canonical contact structure on the space of light rays based on symplectic reduction techniques.

5. The contact structure in $\mathcal{N}$ and symplectic reduction

Finally, in this section, we will illustrate the construction of the contact structure in $\mathcal{N}$ in a equivalent but more elegant way as done in section 4.2.

5.1. The coisotropic reduction of $\mathbb{N}^+$ and the symplectic structure on the space of scaled null geodesics $\mathcal{N}_s$

The celebrated Theorem of Marsden–Weinstein [MW74] claims that a $2m$–dimensional symplectic manifold P, in which a Lie group G acts preserving the symplectic form ω and possessing an equivariant momentum map, can be reduced into another $2(m-r)$–dimensional symplectic manifold P_μ, called the Marsden-Weinstein reduction of P with respect to μ, under the appropriate conditions where μ is an element in the dual of the Lie algebra of the group G and r is the dimension of the coadjoint orbit passing through μ.

The purpose of this section is to show that it is possible to derive the canonical contact structure on the space of light rays by a judiciously use of Marsden-Weinstein reduction when the geodesic flow defines an action of the Abelian group $\mathbb{R}$ in TM. However we will choose a different, simpler, however more general path here. Simpler in the sense that we will not need the full extent of MW reduction theorem, but a simplified version of it obtained when restricted to scalar momentum maps, but more general in the sense that it will not be necessary to assume the existence of a group action. Actually the setting we will be using is a particular instance of the scheme called generalized symplectic reduction (see for instance [Ca14, Ch.7.3] and references therein).

The result we are going to obtain is based on the following elementary algebraic fact. Let (E,ω) be a linear symplectic space. Let $W \subset E$ be a linear subspace. We denote by $W^\perp$ the symplectic orthogonal to W, i.e., $W^\perp = \{u \in E \mid \omega(u,w) =$

$0\,, \forall w \in W\}$. A subspace W is called coisotropic if $W^\perp \subset W$. It is easy to show that for any subspace W, $\dim W + \dim W^\perp = \dim E$. Hence it is obvious that if H is a linear hyperplane, that is a linear subspace of codimension 1, then H is coistropic (clearly because ω_H is degenerate, then $H \cap H^\perp \neq \{\mathbf{0}\}$ and because $H^\perp$ is one-dimensional, then $H^\perp \subset H$). Moreover the quotient space $H/H^\perp$ inherits a canonical symplectic form $\bar\omega$ defined by the expression:

$$\bar\omega(u_1 + H^\perp, u_2 + H^\perp) = \omega(u_1, u_2)\,, \qquad \forall u_1.u_2 \in H\,.$$

The linear result above has a natural geometrical extension:

Theorem 5.1 *Let (P,ω) be symplectic manifold and $i\colon S \to P$ be a hypersurface, i.e., a codimension 1 immersed manifold. Then:*

i. The symplectic form ω induces a 1-dimensional distribution K on S, called the characteristic distribution of ω, defined as $K_x = \ker i^*\omega_x = T_xS^\perp \subset T_xS$.

ii. If we denote by $\mathcal{K}$ the 1-dimensional foliation defined by the distribution K and $\overline{S} = S/\mathcal{K}$ has the structure of a quotient manifold, i.e., the canonical projection map $\rho\colon S \to S/\mathcal{K}$ is a submersion, then there exists a unique symplectic form $\bar\omega$ on $\overline{S}$ such that $\rho^*\bar\omega = i^*\omega$.

iii. If $\omega = -d\theta$ and there exists $\bar\theta$ a 1-form on $\overline{S}$ such that $\rho^*\bar\theta = i^*\theta$, then $\bar\omega = -d\bar\theta$.

Proof. The proof of (i) is just the restriction of the algebraic statements above to $W = T_xS \subset E = T_xP$.

To proof (ii), notice that a vector tangent to the leaves of $\mathcal{K}$ is in the kernel of $i * \omega$, then for any vector field X on S tangent to the leaves of $\mathcal{K}$, i.e, projectable to 0 under ρ, we have $i_X(i^*\omega) = 0$, and $\mathcal{L}_X(i^*\omega) = 0$, then the 2-form $i^*\omega$ is projectable under ρ.

The statement (iii) is trivial because $\rho^*\bar\omega = i^*\omega = i^*(-d\theta) = -di^*\theta = -d\rho^*\bar\theta = \rho^*(-d\bar\theta)$ and ρ is a submersion. $\square$

The previous theorem states that any hypersurface on a symplectic manifold is coisotropic and that, provided that the quotient space is a manifold, the space of leaves of its characteristic foliation, inherits a symplectic structure. Such space of leaves is thus the reduced symplectic manifold we are seeking for and it will be called the coisotropic reduction of the hypersurface S. In addition to the previous reduction mechanism, we will also use the following passing to the quotient mechanism for hyperplane distributions.

Theorem 5.2 *Let $(P,\omega = d\theta)$ be an exact symplectic manifold and $\pi\colon P \to N$ be a submersion on a manifold of dimension* $\dim P - 1$ *and such that it projects the hyperplane distribution $H = \ker\theta$, that is there exists a hyperplane distribution H^N in N such that for any $x \in P$, $\pi_*(x)H_x = H^N_{\pi(x)}$. Then H^N defines a contact structure on N.*

Proof. Notice that necessarily, $\ker \pi_*(x) = H_x^\perp$ and ω induces a symplectic form $\bar{\omega}_x$ in $H/H^\perp$ because Thm. 5.1. Moreover $H_x/H_x^\perp \cong H_{\pi(x)}^N$ and it inherits a symplectic form $\bar{\omega}_x$. Finally, if we pick up a local section σ of the submersion π; then the 1-form $\sigma^*\theta$ is such that $H^N = \ker \sigma^*\theta$ and $d(\sigma^*\theta)$ coincides with the symplectic form $\bar{\omega}_x$ when restricted to H_x^N. □

The two previous results, Thm. 5.1 and 5.2, hold the key to understand how the quotient space $\mathcal{N}$ inherits a canonical contact structure. Consider again a spacetime $(M, \mathbf{g})$ and the canonical identification provided by the metric $\hat{\mathbf{g}} \colon \widehat{T}M \to \widehat{T}^*M$ (which is just the Legendre transform corresponding to the Lagrangian function $L_{\mathbf{g}}(x, v) = \frac{1}{2}\mathbf{g}_x(v, v)$ on TM). As we discussed at the beginning of Sect. 4, Eqs. (4.5), (4.9), we can pull-back the canonical 1-form θ on T^*M along $\hat{\mathbf{g}}$ as well the canonical symplectic structure ω (Sect. 4.1), that is, we obtain:

$$\theta_g = \widehat{\mathbf{g}}^*\theta\,, \qquad \omega_g = \widehat{\mathbf{g}}^*\omega = -d\theta_g\,,$$

and $(\widehat{T}M, \omega_{\mathbf{g}})$ becomes a symplectic manifold. Moreover $\mathbb{N}^+ \subset \widehat{T}M$ defines an hypersurface, hence by Thm. 5.1 we can construct its coisotropic reduction.

We will denote by $\mathcal{N}_s$ the space of equivalence classes of future-oriented null geodesics that differ by a translation of the parameter. Thus two parametrized null geodesics $\gamma_1(t)$, $\gamma_2(t')$ are equivalent if there exists a real number s such that $\gamma_2(t') = \gamma_1(t+s)$. The equivalence class of null geodesics containing the parametrized geodesic $\gamma(t)$ such that $\gamma'(0) = v$ will be denoted by γ_v.

Clearly there is a natural projection $\pi \colon \mathcal{N}_s \to \mathcal{N}$ defined as $\pi(\gamma_v) = [\gamma]$. The space $\mathcal{N}_s$ is sometimes called the space of *scaled null geodesic* and describes equivalence classes of null geodesics distinguishing different scale parametrizations.

Theorem 5.3 *Let $(M, \mathbf{g})$ be a spacetime, then:*

i. The characteristic distribution $K = \ker \omega_{\mathbf{g}} \mid_{\mathbb{N}^+}$ is generated by the restriction of the geodesic spray $X_{\mathbf{g}}$ to $\mathbb{N}^+$ and $\mathbb{N}^+/K$ can be identified naturally with the space of scaled null geodesics $\mathcal{N}_s$.

ii. If M is strongly causal, $\mathcal{N}_s$ is a quotient manifold of $\mathbb{N}^+$, and it becomes a symplectic manifold with the canonical reduced symplectic structure obtained by coisotropic reduction of $\omega_{\mathbf{g}}$.

Proof. To prove [i] we just check that $\omega_{\mathbf{g}_v}(X_{\mathbf{g}}, Y) = dL_v(Y) = Y_v(L) = 0$ for all $Y \in T_v(\mathbb{N}^+)$ because $\mathbb{N}^+ = L^{-1}(\mathbf{0})$ where L is the Lagrangian function $L(u) = \frac{1}{2}\mathbf{g}(u, u)$

Notice that the flow φ_t of the geodesic spray $X_{\mathbf{g}}$ is such that $\varphi_s(\gamma(t)) = \gamma(t+s)$ where $\gamma(t)$ is a parametrized geodesic. Then the quotient $\mathbb{N}^+/K$ corresponds exactly to the notion of scaled null geodesic before. We will denote, as before, by $\rho \colon \mathbb{N}^+ \to \mathcal{N}_s$ the canonical projection and, with the notations above, we get simply that $\rho(v) = \gamma_v$.

As M is strongly causal, the proof of [ii] mimics the proof of Prop. 2.2. Hence because [ii] in Thm. 5.1, we conclude that the quotient manifold inherits a canonical symplectic structure by coisotropic reduction of $\omega_{\mathbf{g}}$. $\square$

References

[Ab87] R. Abraham, J. Marsden: *Foundations of Mechanics*. Addison-Wesley, 1987.

[Ab88] R. Abraham, J. Marsden, T. Ratiu: *Manifolds, tensor analysis, and applications*. Springer-Verlag, 1988.

[Ar89] V. I. Arnold: *Mathematical methods of classical mechanics*.Springer Verlag, 1989.

[Ba14] A. Bautista, A. Ibort, J. Lafuente: *On the space of light rays of a spacetime and a reconstruction theorem by Low*. Cllass. Quantum Grav., **31** (2014) 075020.

[Ba15] A. Bautista, A. Ibort, J. Lafuente: *Causality and skies: is non-refocussing necessary?*. Class. Quantum Grav., **32** (2015) 105002. Doi:10.1088/0264-9381/32/10/105002.

[Ba15b] A. Bautista: *Causality, light rays and skies*. Ph. D. Thesis (2015).

[BE96] J. K. Beem, P. E.Ehrlich, K. L.Easley: *Global Lorentzian Geometry*. Marcel Dekker, 1996.

[Be03] A. N. Bernal, M. Sánchez: *On smooth Cauchy hypersurfaces and Geroch's splitting theorem*. Commun. Math. Phys. 243, 2003, 461–470.

[Br93] G. E. Bredon: *Topology and Geometry*. Springer-Verlag, 1993.

[Bri70] F. Brickell, R. S. Clark: *Differentiable manifolds. An Introduction*. Van Nostrand Reinhold, 1970.

[Ca01] A. Cannas da Silva: *Lectures on symplectic geometry*. Springer-Verlag, 2001.

[Ca14] J. F. Cariñena, A. Ibort, G. Marmo, G. Morandi: *Geometry from dynamics: Classical and Quantum*. Springer-Verlag (2014).

[Ch10] V. Chernov, S. Nemirovski: *Legendrian Links, Causality, and the Low Conjecture*. Geom. Funct. Analysis, **19** (5) 1320–1333 (2010).

[G08] H. Geiges: *An introduction to contact topology*. Cambridge University Press, 2008.

[HE73] S. W. Hawking, G. F. R. Ellis: *The large scale structure of space-time*. Cambridge University Press, Cambridge, 1973.

[KT09] B. Khesin, S. Tabachnikov: *Pseudo-riemannian geodesics and billiards*. Adv. Math. 221, 2009, 1364–1396.

[LM87] P. Libermann, C. M. Marle: *Symplectic geometry and analytical mechanic*. D. Reidel Publishing Company, Dordrecht, 1987.

[Lo88] R. J. Low: *Causal relations and spaces of null geodesics*. PhD Thesis, Oxford University (1988).

[Lo89] R. J. Low: *The geometry of the space of null geodesics*. J. Math. Phys. 30(4) (1989), 809–811.

[Lo90] R. J. Low: *Twistor linking and causal relations.* Classical Quantum Gravity 7 (1990), 177–187.

[Lo90-2] R. J. Low: *Spaces of causal paths and naked singularities.* Classical Quantum Gravity 7 (1990), 943–954.

[Lo93] R. J. Low. *Celestial Spheres, Light Cones and Cuts.* J. Math. Phys. 34 (1993), no. 1, 315–319.

[Lo94] R. J. Low: *Twistor linking and causal relations in exterior Schwarzschild space.* Classical Quantum Gravity 11 (1994), 453–456.

[Lo98] R. J. Low: *Stable singularities of wave-fronts in general relativity.* J. Math. Phys. 39 (1998), 3332–3335.

[Lo01] R. J. Low: *The space of null geodesics.* Proc. Third World Congress of Nonlinear Analysts, Part 5 (Catania, 2000). Nonlinear Anal. 47, 2001, no. 5, 3005–3017.

[Lo06] R. J. Low: *The space of null geodesics (and a new causal boundary).* Lecture Notes in Physics 692, 2006, pp. 35–50.

[MW74] J.E, Marsden, A. Weinstein. *Reduction of symplectic manifolds with symmetry*, Rep. Mathematical Phys., 5, 1, 121–130, (1974).

[Mi08] E. Minguzzi, M. Sánchez: *The causal hierarchy of spacetimes*, Zurich: Eur. Math. Soc. Publ. House, H. Baum, D. Alekseevsky (eds.), Recent developments in pseudo-Riemannian geometry of ESI Lect. Math. Phys., 299–358 (2008). arXiv:gr-qc/0609119.

[Na04] J. Natario, P. Tod: *Linking, Legendrian linking and causality.* Proc. London Math. Soc. (3) **88** 251–272 (2004).

[On83] B. O'Neill: *Semi-Riemannian geometry with applications to Relativity.* Academic Press. New York, 1983.

[Pe72] R. Penrose: *Techniques of Differential Topology in Relativity.* Regional Conference Series in Applied Mathematics, SIAM, 1972.

[Pe79] R. Penrose: *Singularities and time-asymmetry.* General Relativity: An Einstein Centenary (S. W.Hawking & W. Israel, eds.), Cambridge University Press, 1979.

[Wa83] F. W. Warner: *Foundations of differentiable manifolds and Lie groups.* Springer-Verlag, 1983.

Universal groups and super regular tesselations

G. Brumfiel, H. Hilden, M. T. Lozano*, J. M. Montesinos-Amilibia†, E. Ramírez, H. Short, D. Tejada and M. Toro

Stanford University, Stanford, Ca, USA, brumfield@math.stanford.edu

University of Hawaii, Honolulu,Hi, USA, mike@math.hawaii.edu

Universidad de Zaragoza, 50009 Zaragoza, Spain, tlozano@unizar.es

Universidad Complutense, 28049 Madrid, Spain, jose-montesinos@mat.ucm.es

CIMAT, Mexico, Mexico, kikis@cimat.mx

Université de Provence, Marseille, France, hamish.short@cmi.uni-mrs.fr

Universidad Nacional de Colombia, Medellín, Colombia, dtejada@unalmed.edu.co, mmtoro@unalmed.edu.co

Los coautores de José Montesinos le dedicamos nuestra parte del trabajo en su 70 cumpleaños, con afecto.

ABSTRACT

Here we study the tessellation of H^3 by regular right dihedral angled dodecahedra. Any one dodecahedron of the tessellation is a fundamental domain for $\mathbf{U}$, the orbifold group of the Borromean rings with singular angle 90°.

A family of planes of the tessellation is defined in such a way that no two distinct planes in the family intersect. The group $\mathbf{U}$ is universal. This means that given any closed orientable 3-manifold M, there is a finite index subgroup G of $\mathbf{U}$ such that M is homeomorphic to H^3/G.

The group $\mathbf{U}$ does not leave the family of planes invariant but there is an index two subgroup $\mathbf{D}$ of $\mathbf{U}$ that does. The group $\mathbf{D}$ is also universal. The family of planes partitions H^3 into components that are convex manifolds with totally geodesic disconnected boundaries. A consequence of the existence of $\mathbf{D}$ is that any closed orientable 3-manifold can be constructed from the dodecahedra in any one component.

2010 Mathematics Subject Classification: 57M25, 57M12, 57M50.

Key words: Super regular tessellation, orbifold, universal group.

*Author partially supported by project MTM2013-45710-c2 and DGA Grupo Consolidado E15.

†Author partially supported by project MTM2012-30719.

1. Introduction

This paper deals with regular and super regular tesselations of E^3 and H^3. A regular tesselation is a partition by regular polyhedra, cubes, tetrahedra, octahedra, dodecahedra, or icosahedra, such that any two polyhedra intersect in a face or edge or vertex of both or not at all. We say that regular tesselation is *super regular* if it induces regular tesselations in lower dimensions. For example in E^2 the tesselations by squares or equilateral triangles are super regular but the tesselation by regular hexagons is not because the edges do not fit together to give a tesselation of a line.

The only regular tesselation of E^3 is by cubes and it is easily seen to be super regular. Let T_E be the tesselation of E^3 by $2\times2\times2$ cubes with odd integer coordinate vertices from now on. The only super regular tesselation of H^3 is by regular right dihedral angled dodecahedra. Let's call it T_H (See [5] and [3]). Distinct tesselating planes in a super regular tesselation of E^3 or H^3 either intersect at right angles or do not intersect at all.

In this paper we show how to define a large subfamily of planes of the tesselation T_H of H^3 that do not intersect. This family, denoted by X_{ac}, naturally defines via intersection of half spaces a decomposition of H^3 into convex submanifolds with totally geodesic boundary.

There is a group **U**, the orbifold group of the Borromean rings with order 4, for which any one dodecahedron of the tesselation T_H is a fundamental domain. The group **U** is universal. (See [4] and [2]). This means that given any closed orientable 3–manifold M there is a finite index subgroup G of **U** such that the quotient space H^3/G is homeomorphic to M. Thus M has the structure of a hyperbolic orbifold, not manifold, induced by this group G.

The group **U** does not preserve the set of planes X_{ac} and thus does not act on the set of convex manifolds into which H^3 is partitioned.

The main result of this paper, Theorems 3.1 show that **U** has an index two subgroup **D** that does preserve the set of planes X_{ac}, that does act on the set of convex manifolds into which H^3 is partitioned and is itself universal. A consequence of the existence of **D**, Theorem 4.2, is that any closed oriented 3–manifold M can be constructed from the dodecahedra in any one of the convex manifolds into which H^3 is decomposed, in a certain way.

The organization of the paper is as follows: in Section 2, we study several subgroups of **U** defined as kernels of homomorphisms of **U** to abelian groups. And in Section 2 we define the family of planes X_{ac} such that no two distinct planes in X_{ac} intersect. It is the family of planes in X_{ac} that gives rise to the partition of H^3 into convex submanifolds with totally geodesic boundary.

In Section 3 we show that an index two subgroup **D** of **U** is universal. As **D** also acts on the set of planes X_{ac} we obtain the main theorem of this paper, Theorems 3.1. In the final section we show (Theorem 4.2) that any closed orientable 3–manifold can be obtained by pasting together the dodecahedra in one of the convex submanifolds into which H^3 is partitioned by the family of planes in X_{ac}.

2. The family X_{ac} and various subgroups of U

The group **U** is the orbifold group of the orbifold (S^3, B_{444}), orbifold structure in S^3 with the Borromean rings as singular set with angle $90°$, pictured below in Figure 1.

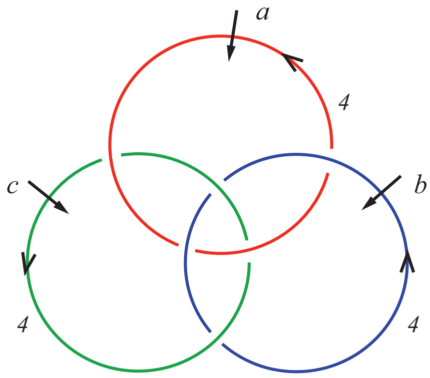


Figure 1: Borromean Rings.

A presentation of **U** is obtained from the Wirtinger presentation of the fundamental group of the Borromean rings by adding the branch relations a^4, b^4, c^4. Here is a presentation of U:

$$\begin{aligned}\mathbf{U} = \langle a, b, c \,|\, a\, b\bar{c}\bar{b}c &= b\bar{c}\bar{b}c\, a,\ a^4, \\ b\, c\bar{a}\bar{c}a &= c\bar{a}\bar{c}a\, b,\ b^4, \\ c\, a\bar{b}\bar{a}b &= a\bar{b}\bar{a}b\, c,\ c^4 \,\rangle\end{aligned} \tag{2.1}$$

The generators a, b, and c arise from the three meridian generators for the three components of the Borromean rings. We now define some homomorphisms of **U** by defining the image of a, b, and c and verifying the relations

$$\begin{gathered}\alpha : \mathbf{U} \longrightarrow Z_4 = \{0, 1, 2, 3; +\} \\ \alpha(a) = \alpha(b) = \alpha(c) = 1 \\ \mathbf{A} := \text{kernel } \alpha\,.\end{gathered} \tag{2.2}$$

As there are nine conjugacy classes of rotations in **U**, represented by a^j, b^j, and c^j; $j = 1, 2, 3$, the group **A** contains no rotations and therefore acts freely on H^3. Similarly

$$\begin{gathered}\delta : U \rightharpoonup Z_2 = \{0, 1; +\} \\ \delta(a) = 1 \quad \delta(b) = \delta(c) = 0 \\ D := \text{kernel } \delta\,.\end{gathered} \tag{2.3}$$

Let $\mathbf{SQ}$ be the subgroup of $\mathbf{U}$ generated by squares. Then $\mathbf{SQ}$ is a normal subgroup and we define the natural homomorphism $\gamma : \mathbf{U} \longrightarrow \mathbf{U}/\mathbf{SQ}$. Every element of $\mathbf{U}/\mathbf{SQ}$ has order two so $\mathbf{U}/\mathbf{SQ}$ is abelian and γ factors through the abelianization of $\mathbf{U}$

$$\gamma : \mathbf{U} \longrightarrow \mathbf{U}/[\mathbf{U},\mathbf{U}] \longrightarrow \mathbf{U}/\mathbf{SQ}\,. \tag{2.4}$$

It is clear from the presentation of $\mathbf{U}$ that $\mathbf{U}/[\mathbf{U},\mathbf{U}] \cong Z_4 \oplus Z_4 \oplus Z_4$ and that $\mathbf{U}/\mathbf{SQ} \cong Z_2 \oplus Z_2 \oplus Z_2$. We write $a \longrightarrow 100$ instead of $a \longrightarrow (1,0,0)$. Thus

$$\gamma(a) = 100, \quad \gamma(b) = 010, \quad \gamma(c) = 001\,. \tag{2.5}$$

As the elements of $\mathbf{A}$ have even length $\gamma(\mathbf{A})$ is an order four subgroup

$$\begin{aligned} \gamma(\mathbf{A}) &= \{000, 011, 101, 110\} \\ \mathbf{SQ} &= \operatorname{kernel}\gamma\,. \end{aligned} \tag{2.6}$$

At this point it is useful to reproduce a figure from ([2]). D_0 is a regular Euclidean and hyperbolic dodecahedron in the Klein model for H^3. Depicted in Figure 2 are D_0 itself, the intersection of D_0 with the positive octant and the axes of rotation for the generators a, b, and c of $\mathbf{U}$, also labelled a, b, and c. The dodecahedron D_0 is a fundamental domain for $\mathbf{U}$.

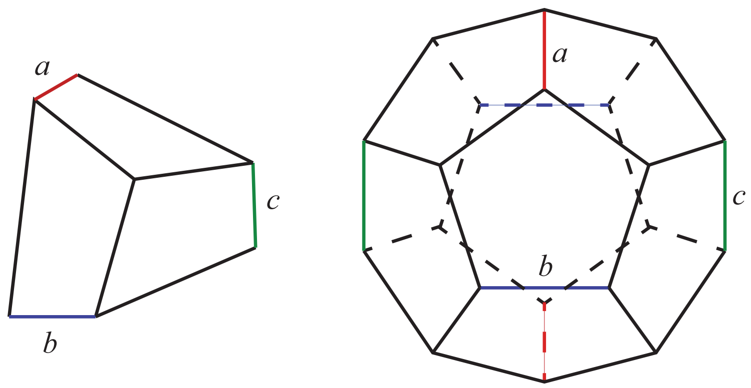


Figure 2: The dodecahedron D_0.

We need to distinguish three types of pentagons in the tesselation T_H, types ac, ba, and cb. We say a pentagon is of type ac, for example, if one of its edges lies on an axis of rotation for a rotation conjugate to a and the vertex opposite this edge lies on an axis of rotation orthogonal to the plane of the pentagon and the corresponding rotation is conjugate to c. As every pentagon is equivalent to a pentagon in D_0 the three types are well defined. In the following we will refer to type a, b, and c

vertices and type a, b, and c edges. There are no type ca, ab, or bc pentagons. A "neighborhood" of a c vertex of an ac pentagon contained in a plane of T_H consists of four ac pentagons. The following proposition follows from this.

Proposition 2.1 *All the pentagons in a plane of T_H are of the same type.* □

Thus we may speak of type ac, ba, and cb planes.

Proposition 2.2 *The groups* $\mathbf{A}$ *and* $\mathbf{U}$ *act transitively on the axes of a given type and on the planes of a given type.*

Proof. Suppose, for example, Q is an ac plane and π is a pentagon in Q belonging to a dodecahedron D. There is an element $u \in \mathbf{U}$ such that $u(D) = D_0$ and $u(\pi)$ lies in one of the four planes intersecting D_0 in its four ac pentagons. The two a rotations about the a axes intersecting D_0 and the b rotation about the b axes intersecting D_0 can be used to send any one of these four planes to any other. Transitivity of $\mathbf{A}$ on ac planes follows by composing an element of $\mathbf{U}$ with some power of a c rotation about a c axis perpendicular to Q. □

If Q is an ac plane and G is a subgroup of $\mathbf{U}$ we let $G_Q = \{g \in G \mid g(Q) = Q\}$. We are particularly interested in $\mathbf{U}_Q$ and $\mathbf{A}_Q$.

Fix Q an ac plane. There are some elements of $\mathbf{U}$ that obviously belong to $\mathbf{U}_Q$.

1. 180° rotations about a axes lying in Q.

2. All rotations about c axes perpendicular to Q.

We define the group REFL_Q to be the group generated by type 1 elements.

The axes of type a lying in Q partition Q into convex sets as follows. If x is a point not lying on any a axis and ℓ is an a–axes, let $H_\ell(x)$ be the closed half plane in Q defined by ℓ that contains x. Then $K(x) = \cap H_\ell(x)$ where the intersection is taken over all a axes ℓ lying in Q.

Let K_0 be a convex component and let G_{K_0} be the group generated by all rotations about c axes that intersect K_0.

Proposition 2.3 *For any $g \in \mathbf{U}_Q$, $g = rv$ where $r \in \mathrm{REFL}_Q$ and $v \in G_{K_0}$.*

Proof. Fix π_0 a pentagon in K_0 and let π_1 be an adjacent pentagon in the sense depicted in Figure 3. That is π_0 and π_1 have edges that intersect an a–axis that we call ℓ and the common edge $e = \pi_0 \cap \pi_1$, is not contained in an axis of rotation.

Let A and B be the c vertices of π_0 and π_1, respectively and let M be the point $\ell \cap e$.

Now, if $g \in \mathbf{U}_Q$, let $\pi_2 = g(\pi_0)$ and let C be the c vertex of π_2. There is a geodesic AC in Q that intersects a finite number of a axes $\ell_1, \ldots, \ell_m$ is the manner depicted in Figure 4.

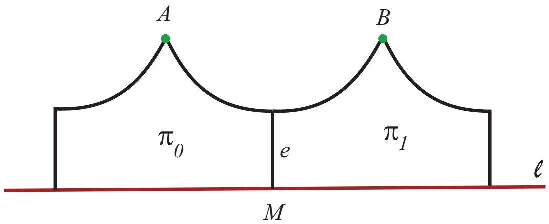


Figure 3: Adjacent pentagons.

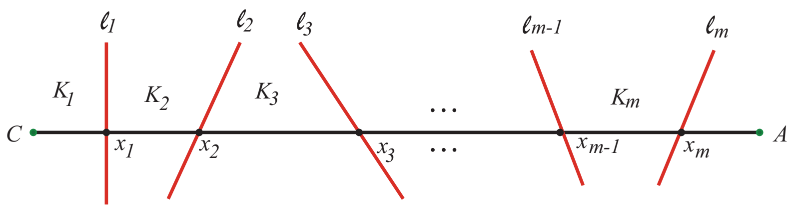


Figure 4: Geodesic AC.

Let $x_\ell = \ell_\ell \cap CA$ and let K_i be the convex component containing the line segment $x_{i-1}x_i$; $2 \leqq i \leqq m$. Let K_1 be the component containing C and K_0, as we know, is the component containing A. Let r_i equal $180°$ rotation in ℓ_i. The restriction of r_i to Q is just a reflection in ℓ_i. Then $r_i(K_i) = K_{i+1}$, $1 \leqq i \leqq m-1$ and $r_m(K_m) = K_0$. If we let $r = r_m r_{m-1} \cdots r_2 r_1$, then $r(\pi_2) \subset K_0$.

Next we define a graph Γ in K_0. The component K_0 is naturally tesselated by octagons formed by joining four pentagons with a common c vertex as depicted in Figure 5.

The vertices of Γ are the c vertices in K_0 and there is an edge between two vertices whenever the corresponding octagons intersect in a common edge. Using the convexity of K_0, or collapsing K_0 onto Γ, we see that Γ is a tree. Let D be the c–vertex of $r(\pi_2) = rg(A)$. As there is an edge from A to B there is a path of even length $V_0V_1V_2 \cdots V_{2n}$ from D to either A or B.

For j odd let c_j be the c rotation conjugate to c whose axis intersects K_0 in V_j. Then for k =1, 2, or 3 $c^k_{2j-1}(V_{2j-2}) = V_{2j}$; $1 \leqq j \leqq n$. Let $\widehat{c_{2j-1}} = c^k_{2j-1}$ where k is such that $c^k_{2j-1}(V_{2j-2}) = V_{2j}$. Now let $h_1 = \widehat{c}_{2n-1} \cdots \widehat{c}_3\widehat{c}_1$. Then $h_1 \in G_{K_0}$ and $h_1(D) =$ either A or B. Let $h_2 \in G_{K_0}$ be a c rotation about A or B such that

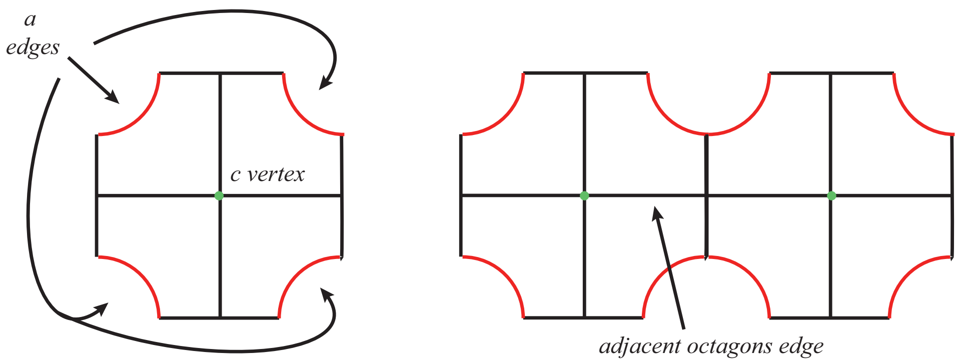


Figure 5: Octagons.

$h_2h_1rg(\pi_0)$ equals either π_0 or π_1. Let $h = h_2h_1$.

We have shown that given $g \in \mathbf{U}_Q$ there exists $r \in \mathrm{REFL}_Q$ and $h \in G_{K_0}$ such that $hrg(\pi_0) = \pi_0$ or π_1.

If $\mu \in \mathbf{U}_Q$ and $\mu(\pi_0) = \pi_0$ its restriction to π_0 is the identity as μ preserves the orientation the a edge of π_0 inherits from ℓ and μ sends the c vertex A to itself. But μ cannot be a reflection in H^3 as $\mathbf{U}$ doesn't contain reflections so that $\mu = id$ on H^3 and $hrg =$ identity implies $g = r^{-1}h^{-1}$. Next we show $\mu(\pi_0) = \pi_1$ cannot occur.

If $\mu(\pi_0) = \pi_1$ then μ must be a hyperbolic translation as it sends ℓ to ℓ and preserves the orientation of ℓ.

Let E_0 and E_1 be the dodecahedra of the tesselation T_H lying "above the page" so that π_0 is a face of E_0 and π_1 is a face of E_1 and so that $E_0 \cap E_1$ is a pentagonal face of both, call it $\widehat{\pi}$. Then $\widehat{\pi}$ lies in a plane orthogonal to Q and $\widehat{\pi} \cap Q$ is the edge e depicted in Figure 3. The a–axis ℓ is orthogonal to $\widehat{\pi}$ so $\widehat{\pi}$ must be a type ba pentagon. There is a type b 90° rotation "about the type b edge or axis" of $\widehat{\pi}$ sending E_0 to E_1. But the element μ also sends E_0 to E_1 as it sends π_0 to π_1. The dodecahedron E_0 is a fundamental domain for $\mathbf{U}$ so that we have shown that μ is both a hyperbolic translation and a 90° rotation which is impossible. $\square$

If $g \in \mathbf{A}_Q$ we see that $g = rv$ where r is a product of 180° rotations in REFL_Q and v is a product of 90° rotations conjugate to c, we write $v = \prod_{i=1}^{n} \widetilde{c}_i$ where $\widetilde{c}_i$ is conjugate to c by writing $\widetilde{c}_i\widetilde{c}_i\widetilde{c}_i$ instead of $\widetilde{c}_i^{-1}$ and $\widetilde{c}_i\widetilde{c}_i$ instead of $\widetilde{c}_i^2$. Applying the homomorphism α, $\alpha(g) = \alpha(r) + \alpha\big(\prod_{i=1}^{n} \widetilde{c}_i\big) = \alpha(r) + n \times \alpha(c)$. But $\alpha(180^\circ \text{ rotation}) = 2$ and $\alpha(c) = 1$ so that $\alpha(g) = 0$ implies n must be even. We summarize this in a proposition.

Proposition 2.4 *For any $g \in \mathbf{A}_Q$, $g = r \prod_{i=1}^{n} \widetilde{c}_i$ where $\widetilde{c}_i$ is a c–rotation conjugate to c, $r \in \mathrm{REFL}_Q$ and n is even.* □

Proposition 2.5 *The homomorphism γ, restricted to $\mathbf{A}_Q$ is trivial.*

Proof. The elements of $\mathbf{U}/\mathbf{SQ}$ have order 2 so if $g = r \prod_{i=1}^{n} \widetilde{c}_i$ with n even as in Proposition 2.4 then $\gamma(g) = \gamma(r) + n\gamma(c) = \gamma(r) + n \cdot 001 = \gamma(r) = 000$ as REFL_Q is generated by 180° rotations which are squares of 90° rotations. □

Now we want to define the type of an *ac* plane which will be an element of the group $\gamma(\mathbf{A}) = \{000, 011, 101, 110\}$. Let P be one of the four planes that intersect D_0 (See Figure 2. In the last section, we shall choose which one of the four.), and let Q be any *ac* plane. By Proposition 2.2 the group $\mathbf{A}$ acts transitively on *ac* planes so there is an element $a \in \mathbf{A}$ such that $a(P) = Q$. Define the type of Q, $t(Q)$, to be $\gamma(a)$.

Proposition 2.6 *The type of Q, $t(Q)$, is well defined.*

Proof. If $a_1(P) = Q$, and $a_2(P) = Q$ then $a_2 a_1^{-1} = \mathbf{A}_Q$ and by Proposition 2.5 $\gamma(a_2 a_1^{-1})$ equals the identity. □

There are exactly three ways that *ac* planes can intersect. If P and Q are *ac* planes then $P \cap Q = P = Q$ or $P \cap Q = \emptyset$ or $P \cap Q$ is an a axis lying in both P and Q. In this last case, we write $P \perp Q$.

Proposition 2.7 *If Q_1 and Q_2 are ac planes with $Q_1 \perp Q_2$ then $t(Q_2) = t(Q_1) + 101$.*

Proof. Suppose $a(P) = Q_1$. Let g be a 90° rotation in the a–axis $Q_1 \cap Q_2$, and let $\widetilde{c}$ be a 90° rotation about any c axis perpendicular to Q_2 such that $\alpha(\widetilde{c}) = (\alpha(g))^{-1}$. Then $\widetilde{c}g \in A$ and $\widetilde{c}ga(P) = Q_2$. Thus $t(Q_2) = \gamma(\widetilde{c}g) + \gamma(a) = 101 + t(Q_1)$. □

We are interested in how a, b, and c act on types of *ac* planes.

Proposition 2.8

$$\begin{aligned} t(cQ) &= t(Q) \\ t(bQ) &= t(Q) + 011 \\ t(aQ) &= t(Q) + 101 \\ t(a^2Q) &= t(Q)\,. \end{aligned} \tag{2.7}$$

Proof. In the first three cases, let $\widetilde{c}$ be a 90° c rotation about a c axis perpendicular to the *ac* planes cQ, bQ, or aQ, respectively and such that $\alpha(\widetilde{c}c) = 0$, $\alpha(\widetilde{c}b) = 0$, or $\alpha(\widetilde{c}a) = 0$, respectively. Then $t(cQ) = t(\widetilde{c}cQ) = 000 + t(Q)$, $t(bQ) = t(\widetilde{c}bQ) =$

$\gamma(\widetilde{c}b) + t(Q) = 011 + t(Q)$ and $t(aQ) = t(\widetilde{c}aQ) = \gamma(\widetilde{c}a) + t(Q) = 101 + t(Q)$. The last case follows immediately from the third case and the fact that the elements of $\mathbf{U}/\mathbf{SQ}$ all have order two. □

Now we define X_{ac} to be the family of ac planes of type 000 or type 011.

Proposition 2.9 *If Q_1 and Q_2 belong to X_{ac} then $Q_1 = Q_2$ or $Q_1 \cap Q_2 = \emptyset$.*

Proof. This follows immediately from Proposition 2.7 and the definition of X_{ac}. □

Theorem 2.1 *The group $\mathbf{D}$, the kernel of the homomorphism $\delta : \mathbf{U} \rightharpoonup Z_2$ leaves the set of planes X_{ac} invariant.*

Proof. As the group $\mathbf{D}$ is generated by a^2, and conjugates of b and c, this follows easily from Proposition 2.8 and the definition of X_{ac}. □

As distinct planes in X_{ac} do not intersect they define a natural partition of H^3 as follows:

If the point x does not belong to any plane in X_{ac} and Q is a plane in X_{ac}, then Q divides H_3 into two closed half spaces, one of which contains x, call it $H_Q(x)$. Let $K(x)$ be the intersection of all the $H_Q(x)$ for $Q \in X_{ac}$. Then $K(x)$ is a convex, (therefore simply connected), manifold with boundary. Two distinct $K(x)$'s either do not intersect or intersect in a plane of X_{ac}, that is, in a boundary plane of both. The component $K(x)$ is a manifold with totally geodesic boundary, albeit with infinitely many boundary components. The component $K(x)$ is naturally tesselated by the dodecahedra it contains.

Referring to Figure 2, we see that D_0 contains four type ac pentagons lying in four ac planes. Using the fact that 90° rotation in a takes one of the "top" ac planes to the other and that 90° rotation in b takes the two "top" ac planes to the two "bottom" ac planes, we see that these four planes represent the four types. Every dodecahedron in T_H contains four ac pentagons lying in four ac planes, one of each type. This is true of any dodecahedron in a $K(x)$ so that two of the four pentagons lie in planes belonging to X_{ac}.

In the next section, we shall show that the group $\mathbf{D}$ of Theorem 2.1 is, like $\mathbf{U}$, universal.

3. The universality of D

The group $\mathbf{D}$ is de orbifold group of the double covering (S^3, Γ) of the orbifold (S^3, B_{444}) branched over one component of the Borromean rings. A proof of the universality of $\mathbf{D}$ can be obtained from the proof of Theorem 1 in [4] observing that the link in Figure 1.16 in [4] is the link Γ. But here we follow here a different approach more convenient for our proposes in next section.

In this proof we need to make minor changes to the proof that $\mathbf{U}$ is universal in [2] and to analyze that proof more carefully.

In [2] we began with an arbitrary closed orientable 3–manifold M and defined the series of branched covering space maps in the third row of (3.1) below.

$$\begin{array}{c} \{e\} \overset{\infty}{\rightarrow}\longrightarrow G_0 \overset{3}{\rightarrow}\longrightarrow G_1 \overset{mn}{\rightarrow}\longrightarrow G_2 \overset{3}{\rightarrow}\longrightarrow G_3 \overset{27}{\rightarrow}\longrightarrow \mathbf{U} \\ H^3 \longrightarrow H^3/G_0 \overset{3-1}{\rightarrow}\longrightarrow H^3/G_1 \overset{mn-1}{\rightarrow}\longrightarrow H^3/G_2 \overset{3}{\rightarrow}\longrightarrow H^3/G_3 \longrightarrow H^3/\mathbf{U} \\ M \longrightarrow S^3 \longrightarrow S^3 \longrightarrow S^3 \longrightarrow S^3 \end{array} \tag{3.1}$$

The branch set in the branched covering $M \approx H^3/G_0 \longrightarrow H^3/\mathbf{U} \approx S^3$ was the Borromean rings (see Figure 1). S^3 has the structure of a hyperbolic orbifold with singular set, the Borromean rings and singular angle 90°. This hyperbolic orbifold structure was pulled back to hyperbolic orbifold structures on M and the three other S^3's and to a hyperbolic manifold structure on H^3. The corresponding orbifold groups and group inclusions and indices are indicated in the first row of (3.1).

In what follows in this chapter, we assume the reader has a copy of [2] available. We introduce an additional Montesinos move to those depicted in [2] in figures one through seven of that paper. This move is displayed below in Figure 6.

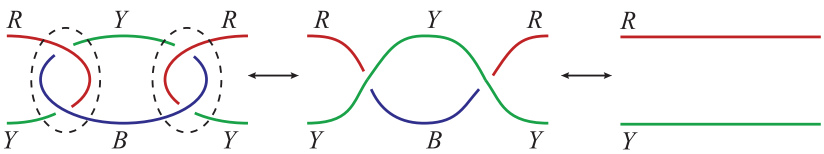


Figure 6: New move.

Using this move we change the left hand side of Figure 11 of [2] to Figure 7 below. We will not need the right hand side of Figure 11 of [2].

This is the first minor change to the proof in [2].

This has the effect of changing the branch set inside the "peanuts" of Figure 12 of [2] so that they look like Figure 7 above. The proof remains the same, the wording verbatim, but now the branch set of Figure 16 of [2] is as in Figure 8 below.

The link L in Figure 8 above can be isotoped to the link of Figure 9.

The link L of Figure 9, like the link of Figure 17 of [2] is universal of type $\{2, 1\}$. This means that any closed orientable 3–manifold is a branched cover of S^3 with this link as branch set and the preimage of any meridian disk consists of a disjoint union of disks each mapped either two to one or one to one to the meridian disk. This link L is the branch set in $S^3 = H^3/G_2$ of (3.1).

Now we proceed, as in [2], to obtain a branched covering of the doubled Borromean rings by composing with a slightly modified rotation $h : S^3 \longrightarrow S^3$. The difference is

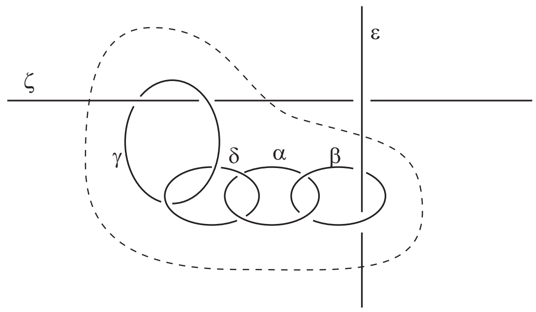


Figure 7: New "peanut".

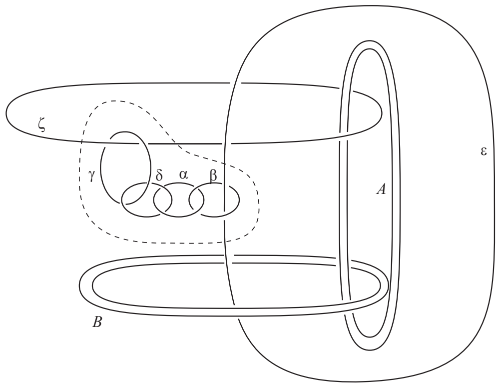


Figure 8: The universal link L.

that this h is four to one whereas the h of [2] was three to one. This is the second minor change.

To see the reason for these minor changes, we must study the twenty–seven to one map $S^3 = H^3/G_3 \longrightarrow H^3/\mathbf{U} = S^3$ and the now four to one map $S^3 = H^3/G_2 \longrightarrow H^3/G_3$ in more detail.

We recall the definition of the twenty–seven to one map defined in [2]. There is a Euclidean crystallographic group $\widehat{\mathbf{U}}$ with fundamental domain the cube $C_0 = \{(x, y, z) \mid -1 \leqq x, y, z \leqq 1\}$ pictured below in Figure 10.

The group $\widehat{\mathbf{U}}$ is generated by 180° rotations in the axes a, b, and c of Figure 10. It has a presentation similar to the presentation of $\mathbf{U}$ in (2.1). The only difference is that the branch relations are a^2, b^2, and c^2 as opposed to a^4, b^4, and c^4. Let T be the homothety of E^3 that changes scale by a factor of 3 so that $T(C_0)$ is the cube

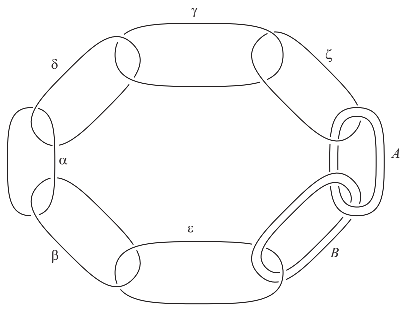


Figure 9: The universal link L.

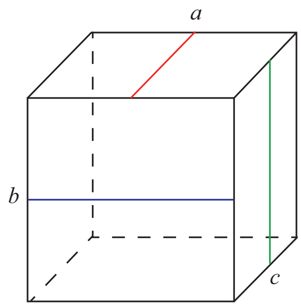


Figure 10: The cube C_0.

$\{(x,y,z) \mid -3 \leqq x,y,z \leqq 3\}$ and let $\widehat{G} = T\widehat{\mathbf{U}}T^{-1}$. Then $\widehat{G} \cong \widehat{\mathbf{U}}$, $\widehat{G}$ is contained in $\widehat{\mathbf{U}}$, and TaT^{-1} is rotation about the axis $(t,0,3)$, $-\infty < t < \infty$, whereas a is a rotation about $(t,0,1)$, $-\infty < t < \infty$. We see that $\widehat{G}$ is an index 27, non–normal subgroup of $\widehat{\mathbf{U}}$, that $E^3/\widehat{G} = S^3$ and that $S^3 = E^3/\widehat{G} \longrightarrow E^3/\widehat{\mathbf{U}} = S^3$ is a twenty–seven to one irregular branched covering of Euclidean orbifolds. In [2] it is shown that the singular set is the Borromean rings and singular angle is 180°.

If we replace the 27 $2 \times 2 \times 2$ cubes that tesselate $T(G_0)$ by 27 dodecahedra isometric to D_0 and we replace C_0 by D_0 and we identify pentagonal faces using hyperbolic isometries instead of Euclidean isometries, we get a twenty–seven to one irregular branched covering of hyperbolic orbifolds $S^3 \longrightarrow S^3$. This is the map $H^3/G_3 \longrightarrow H^3/\mathbf{U}$ of (3.1).

The preimage of the a component of the Borromean rings is a 5–component link pictured in $T(C_0)$ in Figure 11 below. We have labelled the components $a_1, \ldots, a_5$.

Consider the component a_3. It does not belong to an axis of rotation for $\widehat{G}$ so a meridian disc for a_3 is mapped one to one in the map $E^3 \longrightarrow E^3/\widehat{G}$. It does belong

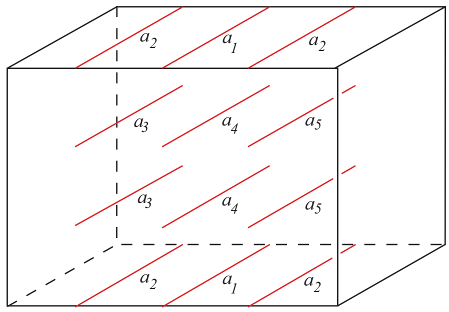


Figure 11: The cube $T(C_0)$

to an axis of rotation for $\widehat{\mathbf{U}}$ so a meridian disc for a_3 is mapped two to one in the map $E^3 \rightharpoonup E^3/\widehat{\mathbf{U}}$. It follows that a meridian disc for a_3 is mapped two to one in the branched covering $S^3 = E^3/\widehat{G} \longrightarrow E^3/\widehat{\mathbf{U}} = S^3$. Similar reasoning shows that meridian discs for a_2, a_4, and a_5 are mapped two to one but meridian discs for a_1 are mapped one to one.

Viewing the twenty–seven to one branched covering as a branched covering of hyperbolic orbifolds, the hyperbolic orbifold structure on $S^3 = H^3/\mathbf{U}$ is pulled back to a hyperbolic orbifold structure on $S^3 = H^3/G_3$. The singular set is a 15 component link $\{a_1, \ldots, a_5, b_2, \ldots, b_5, c_1, \ldots, c_5\}$. The singular angle is $180°$ on all the components except the a_1, b_1, and c_1 components, as the map on meridian discs is two to one. The singular angle is $90°$ on a_1, b_1, and c_1.

The four to one map $H^3/G_2 \longrightarrow H^3/G_3$, which we will shortly describe, will have only type a components in its branch set. The doubled Borromean rings is the six component link $\{a_1, a_2, b_1, b_2, c_1, c_2\}$. The remaining nine components $\{a_3, \ldots, c_5\}$ all have singular angle $180°$. We can now subdivide the axes of rotation in H^3 for $\mathbf{U}$ into fifteen types, labelled by their images in H^3/G_3.

As the nine components $\{a_3, a_4, a_5, b_3, b_4, b_5, c_3, c_4, c_5\}$ all have singular angle $180°$, and do not belong to the branch set for $H^3 \longrightarrow H^3/G_3$, any axis of rotation for $\mathbf{U}$ of one of these nine types must be exactly a 2–fold axis for G_3, i.e., G_3 contains a $180°$ rotation about such an axis but does not contain the $90°$ rotation about that axis. As $G_0 \subset G$ this is true for G_0 also.

Next we consider the 4 to 1 irregular branched covering $S^3 = H^3/G_2 \longrightarrow H^3/G_3 = S^3$ which will be branched over the two a components of the doubled Borromean rings, a trivial link of two components.

Let D be the unit disc pictured in Figure 12 on the left with distinguished points A and B. Then D_1 in the middle is the double branched covering of D with branch point B and $A1$ and $A2$ in D_1 are the two preimages of A. And D_2 on the right is the double branched cover of D_1 with branch point $A1$. The points $\{A21, A22\}$

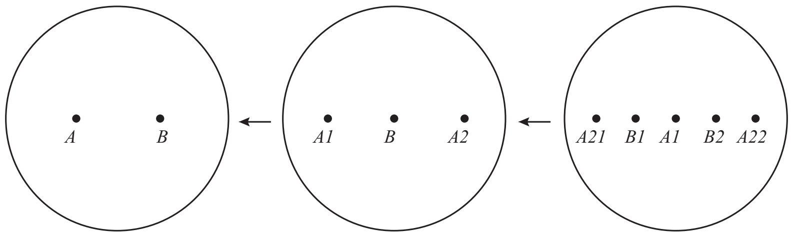


Figure 12: Branched coverings.

and $\{B1, B2\}$ are the preimages of $A2$ and B, respectively. The composite branched covering $D_2 \longrightarrow D$ is an irregular four to one branched covering D to D with branch set $\{A, B\}$. Note that this branched covering when restricted to the boundary is the usual 4–fold covering space map of S^1 to S^1. The preimage of a small disk centered at A is three disks, two of which are mapped in one to one fashion and one of which is mapped two to one. The preimage of a small disk centered at B is two disks both of which are mapped two to one.

Crossing the branched covering $D^2 \longrightarrow D^2$ we have defined with S^1 we get a branched covering $S^1 \times D^2 \longrightarrow S^1 \times D^2$ whose restriction to $S^1 \times S^1$ is equivalent to $\left(e^{i\theta}, e^{i\psi}\right) \longrightarrow \left(e^{4i\theta}, e^{i\psi}\right)$. Displaying S^3 as the union of two solid tori, this branched covering space map extends to a four to one branched covering space map $S^3 \longrightarrow S^3$. The branch set is the trivial link of two components. The preimage of a regular neighborhood of one component is three solid tori, two of which are mapped homeomorphically to the regular neighborhood and one of which is mapped two to one. The preimage of a regular neighborhood of the other component is two solid tori both of which are mapped two to one. This four to one branched covering S^3 to S^3 is the map $H^3/G_2 = S^3 \rightharpoonup S^3 = H^3/G_3$ with branch set the two a components of the doubled Borromean rings, the set $\{a_1, a_2\}$.

We arrange that a_1 is the component whose regular neighborhood has two solid tori both of which are mapped in two to one fashion. Pulling back the h–orbifold structure of $S^3 = H^3/G_3$ to $S^3 = H^3/G_2$ we see that, since the singularity was 90° on the a_1 component it is 180° on both of the preimage components of a_1. Thus, arguing as in the case of the nine components, if ℓ is a type a_1 axis in H^3, the 180° rotation about ℓ belongs to G_2 but the 90° rotation about ℓ does not. As G_0 and G_1 are contained in G_2 this is also true of those groups.

The a_2 component has a singular angle 180° and three components in its preimage, two of these, call them a_{22} and a_{23} inherit singular angles of 180° and in one of them, call it a_{21}, the singularity disappears.

Axes of rotation for $\mathbf{U}$ are labelled by their images in H^3/G_2, H^3/G_3, and $H^3/\mathbf{U}$. The main fact here is that all the type a components in H^3/G_2 do not belong to the branch set of the branched covering $H^3 \longrightarrow H^3/G_2$. So that if an a type axis in H^3

is mapped to a link component a_{jk} in H^3/G_2 the type of axis in H^3 for G_2, 4–fold, 2–fold or not an axis of rotation for G_2 is determined by the singularity angle of a_{jk}, $90°$, $180°$, or not a singularity, respectively. As the axes of type a_1, a_3, a_4, a_5, a_{22}, and a_{23} all correspond to $180°$ singularities, every one of these axes is a 2–fold axis for G_2 and also, as there is no a–branching for $H^3 \longrightarrow H^3/G_0 = M^3 \longrightarrow S^3 = H^3/G_2$, they are also 2–fold axes for G_0 and G_1. All axes of type a_{21} are not axes of rotation for G_2, G_1, or G_0. We summarize this is a proposition.

Proposition 3.1 *Every type a axis of rotation for* $\mathbf{U}$ *is a 2–fold axis of rotation for* G_2*,* G_1 *and for* G_0 *except the type a axes of subtype* a_{21}*. The a axes of subtype* a_{21} *are not axes of rotation for* G_0*,* G_1*, or* G_2*.* □

Corollary 3.1 *The group* G_0 *contains no type a* $90°$ *rotations.* □

As $G_0 \subset G_1$ and H^3/G_1 is S^3 it follows from a theorem of Armstrong (See [1]) that G_1 is generated by rotations. But the group $\mathbf{D}$ contains every type b and every type c rotation and every type a $180°$ rotation. It follows that $\mathbf{D}$ contains G_1 and therefore $\mathbf{D}$ contains G_0. Therefore the group $\mathbf{D}$ is universal.

It was shown in Section 2 that the group $\mathbf{D}$ preserved the set of ac planes X_{ac}. We summarize this in a theorem.

Theorem 3.1 *The index two subgroup* $\mathbf{D}$ *of* $\mathbf{U}$ *is universal and acts on the set of ac planes* X_{ac} *thus preserving the decomposition of* H^3 *into convex manifolds with totally geodesic boundary.* □

In the next section, we make further observations and draw conclusions.

4. Constructing 3–manifolds

In this section we show that every ac plane in X_{ac} contains an a–axis that is the axis for a $180°$ rotation in G_0.

Focusing on the type a axes, we recall that there are five preimages of the a component of the Borromean rings in the covering $S^3 = H^3/G_3 \longrightarrow H^3/\mathbf{U} = S^3$, denoted a_1, a_2, a_3, a_4, a_5 (See Figure 11). This induces a subclassification of the a axes in H^3 according to their image in H^3/G_3. The map from a meridian disk to a_1, a_2, a_3, a_4, or a_5 to a meridian disk for the a component of the Borromean rings is two to one on a_2, a_3, a_4, and a_5 and one to one on a_1. It follows that the singularity in the hyperbolic orbifold H^3/G_2 is $180°$ on the a_2, a_3, a_4, a_5 components and $90°$ on the a_1 component. As a_3, a_4, and a_5 do not belong to the branch set of $H^3/G_0 = M \longrightarrow H^3/G_3 = S^3$ every a_3, a_4, or a_5 axis in H^3 is a 2–fold axis of rotation for G_0, the a_1 and a_2 components are the branch set for the four to one map $S^3 = H^3/G_2 \longrightarrow H^3/G_3 = S^3$. As this map is two to one on a meridian disk to either of the two components in the preimage of the $a1$ component, the singular angle is $180°$ in either of the two a_1 components of the hyperbolic orbifold H^3/G_2. As in the cases of a_3, a_4, and a_5, this implies that every a_1 axis in H^3 is a 2–fold axis for G_0. The a_2 axes remain to be investigated.

The preimage of a_2 in $H^3/G_2 = S^3 \longrightarrow H^3/G_3 = S^3$ consists of three components a_{21}, a_{22}, and a_{23}. Two of these, say a_{22} and a_{23}, have meridian disks that are mapped one to one and the other, a_{21}, has a meridian disk that is mapped two to one. It follows that the singularity is $180°$ on a_{22} and a_{23} in H^3/G_2 and there is no singularity on the a_{21} component. As before a subclassification on type a_2 axes in H^3 is induced, according to their image in H^3/G_2 and axes of type a_{22} and a_{23} are 2–fold axes for G_0. But axes of type a_{21} are not axes of rotation for G_0.

Now, as there is only one a_{21} component in H^3/G_2 the group G_2 acts transitively on the a_{21} components in H^3 and, as $G_2 \subset D$, the group G_2 leaves X_{ac} and X_{ac}^c invariant.

Fix ℓ, an a_{21} axis in H^3. There are two ac planes whose intersection is ℓ. We shall shortly show that one of these planes, call it Q_1, contains an a_1 axis. The other plane is denoted by Q_2.

Recall that there are four choices for the plane P, the four planes that intersect D_0, and that these planes are permuted among themselves by type a rotations and type b rotations. It follows that we may choose P so that the type of Q_1 is 000. We now do this. Then Q_1 belongs to X_{ac} and Q_2 belongs to the complement of X_{ac}.

Let m be any other a_{21} axis in H^3 and let $g(\ell) = m$ for some $g \in G_2$. Let $\widetilde{Q}_1 = g(Q_1)$ and $\widetilde{Q}_2 = g(Q_2)$. Then since g preserves type a_1 axes and type a_{21} axes it follows that $\widetilde{Q}_1$ belongs to X_{ac} and $\widetilde{Q}_1$ contains a type a_1 axis. Thus, first choosing ℓ and then choosing P we obtain the result that every plane in X_{ac} that contains an a_{21} axis also contains an a_1 axis. Next we indicate how to choose ℓ so that Q_1 contains both a_{21} and a_1 axes.

We study the branched covering $S^3 = H^3/G_3 \longrightarrow H^3/\mathbf{U}$. We use Euclidean coordinates for the 27 dodecahedra into which H^3/G_3 is tesselated. Thus these 27 dodecahedra correspond to the 27 $2\times2\times2$ cubes with odd integer vertex coordinates into which the fundamental domain for $\widehat{G}$ is partitioned, $= \{(x,y,z) \mid -3 \leqq x,y,z \leqq 3\}$. Below in Figure 13 we have pictured $\{(x,y,z) \mid -1 \leqq x \leqq 1 \text{ and } -3 \leqq y,z \leqq 3\}$. Let $\beta(t) = \left(\frac{1}{2},t,3\right) = \left(\frac{1}{2},-t,3\right)$; $-2 \leqq t \leqq 0$ and let $D = \left\{(x,y,z) \mid -2\frac{1}{10} \leqq y \leqq 2\frac{1}{10}, x = \frac{1}{2}, 2\frac{9}{10} \leqq z \leqq 3\right\}$.

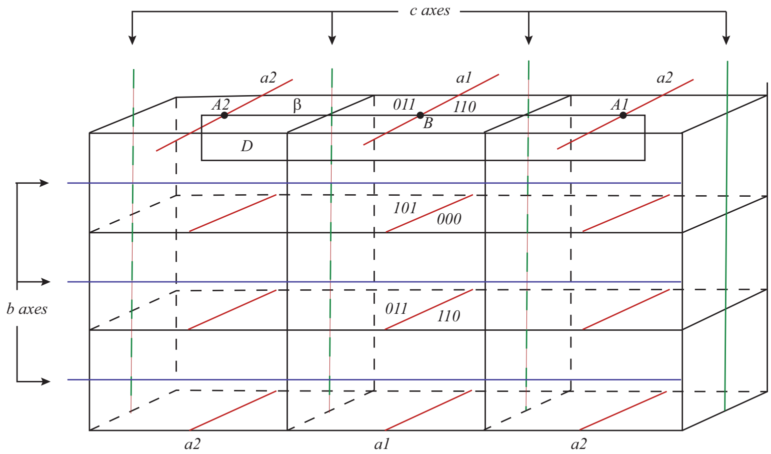


Figure 13: 9 $2 \times 2 \times 2$ cubes.

The first step in constructing the four to one branched covering $H^3/G_2 = S^3 \longrightarrow S^3 = H^3/G_3$ is to construct the two to one covering of H^3/G_3 with branch set a_1. The component a_1 is the trivial knot in $S^3 = H^3/G_3$. The double branched covering can be obtained from the union of $\{(x,y,z) \mid -3 \leqq x,y,z \leqq 3\}$ and its image under $180°$ rotation in the axis $(t,0,3)$, that is to say the set $\{(x,y,z) \mid -3 \leqq x,y,z \leqq 9\}$ with the appropriate identifications on the boundary. The group corresponding to this double branched covering is $G_3 \cap \mathbf{D}$. The restriction of this double branched covering to D and its preimage is illustrated in Figure 14, together with singularity angle and the arc β. The points $A1$ and $A2$ are sent to A and B is sent to B under

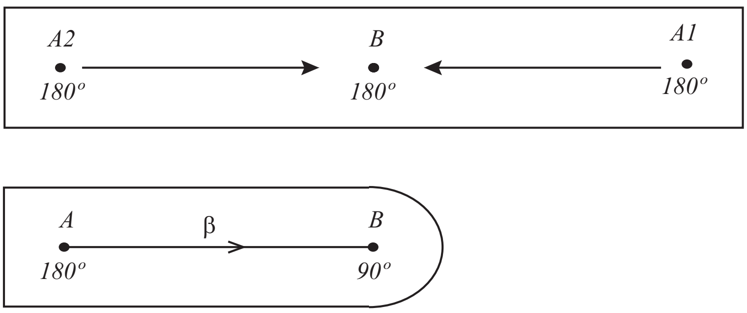


Figure 14: Doble branch covering.

the two to one branched covering map. The two to one branched covering H^3/G_2 to

$H^3/G_3 \cap \mathbf{D}$ will be branched over the $A1$ component. Thus the lift of the arc β to H^3 lies in a plane of the tesselation and connects an a_1 axis to an a_{21} axis. We let ℓ be this a_{21} axis. So now we have the following proposition.

Proposition 4.1 *Every plane in X_{ac} contains a 2–fold axis of rotation for the group G_0.* □

The set of planes X_{ac} decomposes H^3 into components that are convex manifolds with boundary, each component having infinitely many boundary components. Let $K^\star$ be one such component. If D is any dodecahedron in T_H, let A be its center and let B be the center of any dodecahedron in K_0. Then the geodesic AB intersects the set of planes X_{ac} in finitely many points $x_i = Q_i \cap AB$, $1 \leqq i \leqq m$, where the points x_i are in order on $[A, B]$. Let $x_0 = A$ and $x_{M+1} = B$ and let K_i be the component containing the segment $[x_{i-1}, x_i]$. (See Figure 15.)

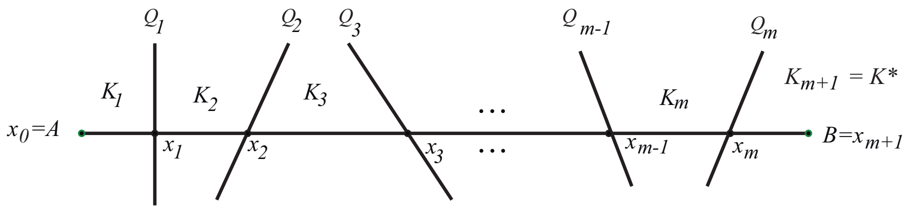


Figure 15: The geodesic AB.

Then if r_j is any 2–fold rotation in plane Q_j we have that $r_j(K_j) = K_{j+1}$ and the product $r_m \cdots r_2 r_1$ sends K_1 to $K^\star$, and D to a dodecahedron in $K^\star$. We summarize this in a theorem.

Theorem 4.1 *Given any dodecahedron D in T_H, there is an element $g \in G_0$ such that g is a product of type a $180°$ rotations in G_0 and $g(D)$ belongs to $K^\star$.* □

Corollary 4.1 *Given a closed orientable 3-manifold M there exits a finite index subgroup G in $\mathbf{D}$, with a fundamental domain contained in any one component K, such that $H^3/G = M$.* □

The idea here is that any closed oriented 3–manifold can be constructed from the dodecahedra in a convex component $K^\star$. The components such as $K^\star$ have very special structure. They are tesselated by dodecahedra, each dodecahedron having two pentagonal faces in the boundary. They are convex, and therefore simply connected. They are naturally cubulated spaces.

References

[1] M. A. Armstrong: *The fundamental group of the orbit space of a discontinuous group*. Math. Proc. Cambridge Phil. Soc. **64** (1968) 299–301.

[2] G. Brumfiel, H. Hilden, M. T. Lozano, J. M. Montesinos, E. Ramirez, H. Short, D. Tejada, M. Toro: *Three manifolds as Geometric branched coverings of the three sphere*. Bulletin of the Mexican Mathematical Society **14(2)** (2008) 263–282.

[3] G. Brumfiel, H. Hilden, M. T. Lozano, J. M. Montesinos, E. Ramirez, H. Short, D. Tejada, M. Toro: *On finite index subgroups of a universal group*. Bulletin of the Mexican Mathematical Society **14(2)** (2008) 283–301.

[4] H. Hilden, M. T. Lozano, J. M. Montesinos, W. Whitten: *On universal groups and three-manifolds*. Invent. Math **87** (1987) 441–456.

[5] W. P. Thurston: *Three-dimensional geometry and topology*. Vol. 1, Edited by Silvio Levy, Princeton Mahematical Series, 35, Princeton University Press, Princeton, NJ, (1997).

Abelian groups of automorphisms of orientable bordered Klein surfaces of topological genus 2

Emilio Bujalance[*], **José Javier Etayo**[†] and **Ernesto Martínez**[‡]

Departamento de Matemáticas Fundamentales
Facultad de Ciencias, UNED
28040 Madrid, Spain
ebujalance@mat.uned.es

Departamento de Álgebra
Facultad de Matemáticas, Universidad Complutense
28040 Madrid, Spain
jetayo@mat.ucm.es

Departamento de Matemáticas Fundamentales
Facultad de Ciencias, UNED
28040 Madrid, Spain
emartinez@mat.uned.es

To Professor José María Montesinos on occasion of his jubilee as an expression of our friendship.

ABSTRACT

In this paper, we obtain the Abelian groups of automorphisms of orientable bordered Klein surfaces of topological genus 2. For each of those groups G we determine the values of k such that G acts on a surface with k boundary components.

2010 Mathematics Subject Classification: 57M60; 20F05, 20H10, 30F50
Key words: Klein bordered surfaces, NEC groups, groups of automorphisms.

[*]Partially supported by MTM2011-23092 and MTM 2014-55812
[†]Partially supported by UCM910444, MTM2011-22435 and MTM2014-55565
[‡]Partially supported by MTM2011-23092 and MTM 2014-55812

1. Introduction and Preliminaries

A natural extension of the definition of compact Riemann surfaces, a concept that implies orientable and unbordered surfaces, is to allow dianalytic transition functions. This leads to surfaces that may be bordered and, or, non-orientable, endowed with a dianalytic structure. These surfaces are called Klein surfaces.

In this work we consider surfaces with non-empty boundary. These surfaces were already considered by F. Klein. For modern references on them, see [1] and [5]. Klein surfaces are determined topologically by three data, the topological genus g, the number of boundary components $k > 0$, and $\alpha = 2$ or 1, according to whether the surface is orientable or not. Then $p = \alpha g + k - 1$ is the algebraic genus of the surface. If $p \geq 2$ the automorphism group of the surface is finite, and its order is bounded above by $12(p-1)$.

A major problem is to determine the groups acting on surfaces of a given topological type. This has been solved for orientable surfaces of topological genus 0 in [3] and for surfaces with connected boundary, that is to say $k = 1$, in Chapter 5 of [5]. We deal here with the next feasible step, which are the orientable surfaces of topological genus $g = 2$. Since there are a lot of groups acting on them, and for such a given group G we need to determine the values of k such that G acts on a surface with k boundary components, we restrict ourselves here to Abelian groups and refer the non-Abelian groups to a sequel of this paper.

In the study of Klein surfaces and their automorphism groups the non-Euclidean crystallographic groups (NEC groups in short) play an essential role. An NEC group Γ is a discrete subgroup of $\mathcal{G}$ (the full group of isometries of the hyperbolic plane $\mathcal{H}$) with compact quotient $\mathcal{H}/\Gamma$. For a Klein surface X with $p \geq 2$ there exists an NEC group Γ, such that $X = \mathcal{H}/\Gamma$, [12].

A finite group G of order N is an automorphism group of a Klein surface $X = \mathcal{H}/\Gamma$ if and only if there exists an NEC group Λ such that Γ is a normal subgroup of Λ with index N and $G = \Lambda/\Gamma$. If G is a finite group there exists a bordered Klein surface X such that G is an automorphism group of X, [11].

First, we give some preliminaries about NEC groups and Klein surfaces.

An NEC group Γ is a discrete subgroup of isometries of the hyperbolic plane $\mathcal{H}$, including orientation-reversing elements, with compact quotient $X = \mathcal{H}/\Gamma$. Each NEC group Γ has associated a signature [10]:

$$\sigma(\Gamma) = (g, \pm, [m_1, \ldots, m_r], \{(n_{i,1}, \ldots, n_{i,s_i}), i = 1, \ldots k\}), \tag{1.1}$$

where g, k, r, m_i, $n_{i,j}$ are integers verifying g, k, $r \geq 0$, m_i, $n_{i,j} \geq 2$. The numbers m_i are the *proper periods*. The brackets $(n_{i,1}, \ldots, n_{i,s_i})$ are the *period-cycles*. Numbers $n_{i,j}$ are the periods of the period-cycle $(n_{i,1}, \ldots, n_{i,s_i})$, also called *link-periods*. We will denote by $[-]$, $(-)$ and $\{-\}$ the cases when $r = 0$, $s_i = 0$ and $k = 0$, respectively. When a proper period or link-period t is repeated r_t times we will write t^{r_t}. Analogously $(-)^s$ means s empty period-cycles.

The signature determines a presentation of Γ, [13], given by the following generators:

x_i $\quad i = 1, \ldots, r$;
e_i $\quad i = 1, \ldots, k$;
$c_{i,j}$ $\quad i = 1, \ldots, k$; $\quad j = 0, \ldots, s_i$;
a_i, b_i $\quad i = 1, \ldots, g$; $\quad$ (if σ has sign " $+$ ");
d_i $\quad i = 1, \ldots, g$; $\quad$ (if σ has sign " $-$ ").

submitted to the following relations:

$x_i^{m_i} = 1$; $\quad i = 1, \ldots, r$;
$c_{i,j-1}^2 = c_{i,j}^2 = (c_{i,j-1}c_{i,j})^{n_{i,j}} = 1$ $\quad i = 1, \ldots, k$; $j = 1, \ldots, s_i$;
$e_i^{-1}c_{i,0}e_ic_{i,s_i} = 1$; $\quad i = 1, \ldots, k$;
$\prod_{i=1}^r x_i \prod_{i=1}^k e_i \prod_{i=1}^g (a_ib_ia_i^{-1}b_i^{-1}) = 1$; $\quad$ (if σ has sign " $+$ ");
$\prod_{i=1}^r x_i \prod_{i=1}^k e_i \prod_{i=1}^g d_i^2 = 1$; $\quad$ (if σ has sign " $-$ ").

The isometries x_i are elliptic, e_i, a_i, b_i are hyperbolic, $c_{i,j}$ are reflections and d_i are glide-reflections.

Every NEC group Γ with signature (1.1) has associated a fundamental region whose area $\mu(\Gamma)$, called *area of the group*, is

$$\mu(\Gamma) = 2\pi \left(\alpha g + k - 2 + \sum_{i=1}^{r} \left(1 - \frac{1}{m_i}\right) + \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{s_i} \left(1 - \frac{1}{n_{i,j}}\right) \right), \tag{1.2}$$

with $\alpha = 2$ or 1 according to the sign be "+" or "-". An NEC group with signature (1.1) actually exists if and only if the right hand side of (1.2) is greater than 0.

We denote by $|\Gamma|$ the expression $\mu(\Gamma)/2\pi$ and call it the *reduced area* of Γ.

If Γ is a subgroup of an NEC group Γ' of finite index N, then also Γ is an NEC group and the following Riemann-Hurwitz formula holds:

$$\mu(\Gamma) = N\mu(\Gamma'). \tag{1.3}$$

Let X be a Klein surface of topological genus g with $k > 0$ boundary components. Then by [12] there exists an NEC group Γ with signature

$$\sigma(\Gamma) = (g, \pm, [-], \{(-), \overset{k}{\ldots}, (-)\}),$$

such that $X = \mathcal{H}/\Gamma$. The sign is "+" or "-" according to X be orientable or not. An NEC group with this signature is called a *bordered surface group*.

For each automorphism group G of a surface $X = \mathcal{H}/\Gamma$ of algebraic genus $p \geq 2$ there exists an NEC group Λ such that $G = \Lambda/\Gamma$ where $\Gamma \subset \Lambda \subset N_{\mathcal{G}}$, [11], and $N_{\mathcal{G}}$ denotes the normalizer of Γ in the group $\mathcal{G}$, the full group of isometries of $\mathcal{H}$.

We recall that the signature of Λ must contain an empty period-cycle or a period-cycle with two consecutive link-periods equal to 2, see [6].

2. The groups of automorphisms

In this paper we obtain the Abelian groups of automorphisms of compact orientable Klein surfaces of topological genus 2. The corresponding result for genus 0 was obtained by Bujalance in [3]. In that work he made use of the following result which relates bordered and unbordered surfaces, see [9, Theorem D].

Proposition 2.1 *Let X be a bordered Klein surface. Then it is possible to embed X in a Klein surface X^* without boundary of the same topological genus, so that every automorphism of X extends to an automorphism of X^*.*

By using this result, our starting point is the list of orientation-preserving automorphism groups of compact Riemann surfaces of genus 2 which was obtained by Broughton in [2]. The list of Abelian groups is the following: C_2, C_3, C_4, $C_2 \times C_2$, C_5, C_6, C_8, C_{10} and $C_2 \times C_6$, where C_n denotes the cyclic group or order n.

The second step is to consider automorphism groups including orientation-reversing involutions, called symmetries. Bujalance and Singerman obtained in [7] the full automorphism group of all Riemann surfaces of genus 2 admitting symmetries. Their list provides new groups to be considered. Precisely, the groups $C_2 \times C_2 \times C_2$, and besides $C_2 \times C_4$ and C_{12}, which are subgroups respectively of $C_2 \times D_4$ and D_{12}.

Finally, it is necessary to consider groups including orientation-reversing automorphisms but not symmetries. However, the unique such group is C_4, see [4], and so no new group appears.

In the following sections we consider each of these twelve groups and we obtain for which values of k they act on an orientable Klein surface of topological genus 2 with k boundary components.

3. Cyclic groups

In this Section we consider the cyclic groups $C_2, C_3, C_4, C_5, C_6, C_8$, C_{10} and C_{12}. For each of them we call X the generator.

The notation throughout the paper will be established here. The main tool for the study of cyclic groups is Theorem 2.4.4 in [5], which was obtained in [8].

We partially restate it as follows.

Theorem 3.1 *Let Γ be a bordered surface NEC group and let N be an even integer. If $\sigma(\Gamma) = (g, \pm, [-], \{(-)^k\})$ and Γ' is another NEC group containing Γ as a normal subgroup with cyclic factor of order N then for some $s' \leq k'$ and some positive even integers $r'_{s'+1}, \ldots, r'_{k'}$, the signature of Γ' is*

$$\sigma(\Gamma') = (g', \pm, [m_1, \ldots, m_r], \{(-)^{s'}, (2^{r'_i})_{i=s'+1,\ldots,k'}\}).$$

Let $c_{i,0} \in \Gamma$ for $1 \leq i \leq p'$ and $c_{i,0} \notin \Gamma$ for $p'+1 \leq i \leq s'$. Let l_i be the order of $\Gamma e_i \in \Gamma'/\Gamma$ for $i \leq p'$, then *i) Each* m_j *divides* N, *for* $j = 1, \ldots, r$.

ii) $k = N \sum_{i=1}^{p'} \frac{1}{l_i} + \frac{N}{2} \sum_{s'+1}^{k'} \frac{r'_i}{2}$.

In order to simplify expressions we fix now some notations. We call $\{\tau_1, \ldots, \tau_\lambda\}$ the increasing ordered set of orders of the elements of the group C_N which are obviously the divisors of N. We split then the k' period-cycles of Γ' as follows. There are s' empty period-cycles and h non-empty period-cycles. Now

$$s' = s_{\tau_1} + s_{\tau_2} + \cdots + s_{\tau_\lambda} + s_0,$$

where s_{τ_i} is the number of period-cycles for which the reflection belongs to Γ and the class of the hyperbolic generator has order N/τ_i in Γ'/Γ, and s_0 is the number of period-cycles such that the corresponding reflection does not belong to Γ.

In case N is odd the situation is simpler, because both h and s_0 are equal to 0 and besides the signs in the signatures of Γ and Γ' are the same.

So that we start with the study of cyclic groups C_3 and C_5.

Group C_3

Since we are dealing with orientable surfaces of topological genus 2, from now on the signature of Γ will be $\sigma(\Gamma) = (2, +, [-], \{(-)^k\})$. According to the notation fixed above let

$$\sigma(\Gamma') = (g', +, [3^{r_3}], \{(-)^{s_1}, (-)^{s_3}\}).$$

Then $k = s_1 + 3s_3$, and so applying (1.3) we have

$$k + 2 = 3\left(2g' + s_1 + s_3 - 2 + \frac{2}{3} r_3\right)$$

so,

$$s_1 + 3s_3 + 2 = 6g' + 3s_1 + 3s_3 - 6 + 2r_3$$

and finally we have

$$8 = 6g' + 2s_1 + 2r_3.$$

We are going to study the different solutions of this equality. First of all, let us observe that g' must be 0 or 1.

If $g' = 0$ then $s_1 + r_3 = 4$, and if $g' = 1$ then $s_1 + r_3 = 1$. This gives the seven kinds of signatures for Γ' which are in the following list:

a1	$(0, +, [3^4], \{(-)^{s_3}\})$	$k = 3s_3$
a2	$(0, +, [3^3], \{(-)^{s_1=1}, (-)^{s_3}\})$	$k = 3s_3 + 1$
a3	$(0, +, [3^2], \{(-)^{s_1=2}, (-)^{s_3}\})$	$k = 3s_3 + 2$
a4	$(0, +, [3], \{(-)^{s_1=3}, (-)^{s_3}\})$	$k = 3s_3 + 3$
a5	$(0, +, [-], \{(-)^{s_1=4}, (-)^{s_3}\})$	$k = 3s_3 + 4$
b1	$(1, +, [3], \{(-)^{s_3}\})$	$k = 3s_3$
b2	$(1, +, [-], \{(-)^{s_1=1}, (-)^{s_3}\})$	$k = 3s_3 + 1$

For each signature we need obtain an epimorphism $\theta : \Gamma' \to C_3$ with kernel Γ. In any case the image of an elliptic element of order 3 must have order 3, the hyperbolic elements corresponding to the s_1 first period-cycles have images of order 3 and the remaining s_3 have an image of order 1, and finally all the reflections belong to Γ.

In order to get all values of k, we limit ourselves to signatures a1, a2 and a3. Consider first the signature a1. Let $\theta_1 : \Gamma' \to C_3$ be defined by

$$\begin{aligned}
\theta_1(x_1) &= X,\ \theta_1(x_2) = X^2,\ \theta_1(x_3) = X,\ \theta_1(x_4) = X^2 \\
\theta_1(e_i) &= 1,\ i = 1, \ldots, s_3 \\
\theta_1(c_{i,0}) &= 1,\ i = 1, \ldots, s_3
\end{aligned}$$

Now we must prove the relations are fulfilled. In this case, the only important point is to check that $\theta(\prod_{i=1}^{4} x_i) = 1$. Because of this reason, we have made the choice of the images of the four elliptic elements x_i.

We will express in short this epimorphism by

$$\theta_1 : (x_1, x_2, x_3, x_4) \to (X, X^2, X, X^2)$$

what means that all canonical generators of Γ' which are omitted, are mapped into 1. We follow this convention throughout the paper.

Now for signatures a2 and a3, it suffices to define respectively

$$\theta_2 : (x_1, x_2, x_3, e_1) \to (X, X^2, X, X^2)$$
$$\theta_3 : (x_1, x_2, e_1, e_2) \to (X, X^2, X, X^2)$$

Group C_5

We proceed as in the group C_3. Now

$$\sigma(\Gamma') = (g', +, [5^{r_5}], \{(-)^{s_1}, (-)^{s_5}\}).$$

Then $k = s_1 + 5s_5$ and applying (1.3) we obtain

$$12 = 10g' + 4s_1 + 4r_5.$$

Then $g' = 0$ and $s_1 + r_5 = 3$. The possible signatures are

a1	$(0,+,[5^3],\{(-)^{s_5}\})$	$k=5s_5$
a2	$(0,+,[5^2],\{(-)^{s_1=1},(-)^{s_5}\})$	$k=5s_5+1$
a3	$(0,+,[5],\{(-)^{s_1=2},(-)^{s_5}\})$	$k=5s_5+2$
a4	$(0,+,[-],\{(-)^{s_1=3},(-)^{s_5}\})$	$k=5s_5+3$

In a similar way to the case C_3 we define the respective epimorphisms by

$$\theta_1:(x_1,x_2,x_3)\to(X,X^2,X^2)$$
$$\theta_2:(x_1,x_2,e_1)\to(X,X^2,X^2)$$
$$\theta_3:(x_1,e_1,e_2)\to(X,X^2,X^2)$$
$$\theta_4:(e_1,e_2,e_3)\to(X,X^2,X^2)$$

The number k attains all values congruent to $0,1,2$ or $3 \pmod 5$.

The next step is the study of cyclic groups of even order. Observe that now the sign of the signature $\sigma(\Gamma')$ can be '+' or '-', and s_0 and h need not to be 0.

Group C_2

The signature of Γ' has the form

$$\sigma(\Gamma')=(g',\pm,[2^{r_2}],\{(-)^{s_1},(-)^{s_2},(-)^{s_0},(2^{t_i})_{i=1,\dots,h}\}).$$

Then $k=s_1+2s_2+\frac{1}{2}\sum_{i=1}^{h}t_i$. Applying (1.3) we obtain

$$6=2\alpha g'+s_1+r_2+2s_0+2h$$

where we recall $\alpha=2$ corresponds to sign '+' in $\sigma(\Gamma')$ and $\alpha=1$ otherwise.

There appear a lot of solutions for this equality. We are going to exhibit only two signatures, which provide all values of k.

a1	$(1,+,[2^2],\{(-)^{s_2}\})$	$k=2s_2$
a2	$(1,+,[2],\{(-),(-)^{s_2}\})$	$k=2s_2+1$

We define the respective epimorphisms θ_1 and θ_2 by means of

$$\theta_1:(a_1,b_1,x_1,x_2)\to(X,X,X,X)$$
$$\theta_2:(a_1,b_1,x_1,e_1)\to(X,X,X,X)$$

Group C_4

The signature of Γ' has the form

$$\sigma(\Gamma')=(g',\pm,[2^{r_2},4^{r_4}],\{(-)^{s_1},(-)^{s_2},(-)^{s_4},(-)^{s_0},(2^{t_i})_{i=1,\dots,h}\}).$$

Then $k = s_1 + 2s_2 + 4s_4 + \sum_{i=1}^{h} t_i$. Applying (1.3) we have

$$10 = 4\alpha g' + 2(r_2 + s_2) + 3(r_4 + s_1) + 4(s_0 + h).$$

Suppose $\alpha = 2$. Recall that if $s_0 \neq 0$ or $h \neq 0$, there exists a reflection c whose image is X^2. On the other hand there must exist an elliptic or hyperbolic element E whose image is X. Then the non-orientable element E^2c must belong to Γ in contradiction with the structure of Γ, see [5, Theorem 2.1.3]. So that we restrict ourselves to the case $s_0 = h = 0$. This restriction will be followed also in the groups C_8 and C_{12}.

Anyway, there are a lot of solutions of the equality which provide all values of k and so we exhibit only four signatures which are sufficient for our purposes.

As we made for C_2 we have

a1	$(0, +, [2^2, 4^2], \{(-)^{s_4}\})$	$k = 4s_4$
a2	$(0, +, [2^2, 4], \{(-)^{s_1=1}, (-)^{s_4}\})$	$k = 4s_4 + 1$
a3	$(0, +, [2, 4^2], \{(-)^{s_2=1}, (-)^{s_4}\})$	$k = 4s_4 + 2$
a4	$(0, +, [2, 4], \{(-)^{s_1=1}, (-)^{s_2=1}, (-)^{s_4}\})$	$k = 4s_4 + 3$

We are going to construct the respective epimorphisms $\theta_1, \theta_2, \theta_3$ and θ_4. In all cases it is straightforward to check the relations are satisfied.

$$\theta_1 : (x_1, x_2, x_3, x_4) \to (X^2, X^2, X, X^3)$$
$$\theta_2 : (x_1, x_2, x_3, e_1) \to (X^2, X^2, X, X^3)$$
$$\theta_3 : (x_1, x_2, x_3, e_1) \to (X^2, X, X^3, X^2)$$
$$\theta_4 : (x_1, x_2, e_1, e_2) \to (X^2, X, X^3, X^2)$$

Group C_6

For the group C_6 we start by taking $s_0 = h = 0$. Then the signature $\sigma(\Gamma')$ is

$$(g', \pm, [2^{r_2}, 3^{r_3}, 6^{r_6}], \{(-)^{s_1}, (-)^{s_2}, (-)^{s_3}, (-)^{s_6})\}).$$

Then $k = s_1 + 2s_2 + 3s_3 + 6s_6$. By (1.3) we have

$$14 = 6\alpha g' + 3(r_2 + s_3) + 4(r_3 + s_2) + 5(r_6 + s_1).$$

As we made before, we are going to show six signatures which provide all values of k:

a1	$(0, +, [3, 6], \{(-)^{s_1=1}, (-)^{s_6}\})$	$k = 6s_6 + 1$
a2	$(0, +, [2^2, 3], \{(-)^{s_2=1}, (-)^{s_6}\})$	$k = 6s_6 + 2$
a3	$(0, +, [2, 3^2], \{(-)^{s_3=1}, (-)^{s_6}\})$	$k = 6s_6 + 3$
a4	$(0, +, [2^2], \{(-)^{s_2=2}, (-)^{s_6}\})$	$k = 6s_6 + 4$
a5	$(0, +, [2, 3], \{(-)^{s_2=1}, (-)^{s_3=1}, (-)^{s_6}\})$	$k = 6s_6 + 5$
a6	$(0, +, [3^2], \{(-)^{s_3=2}, (-)^{s_6}\})$	$k = 6s_6 + 6$

The epimorphisms $\theta_1, \ldots, \theta_6$ are defined below. As in previous cases all relations are evidently satisfied.

$$\begin{aligned}
\theta_1 &: (x_1, x_2, e_1) \to (X^4, X, X) \\
\theta_2 &: (x_1, x_2, x_3, e_1) \to (X^3, X^3, X^2, X^4) \\
\theta_3 &: (x_1, x_2, x_3, e_1) \to (X^3, X^2, X^4, X^3) \\
\theta_4 &: (x_1, x_2, e_1, e_2) \to (X^3, X^3, X^2, X^4) \\
\theta_5 &: (x_1, x_2, e_1, e_2) \to (X^3, X^2, X^4, X^3) \\
\theta_6 &: (x_1, x_2, e_1, e_2) \to (X^2, X^4, X^3, X^3)
\end{aligned}$$

Group C_8

Recall that for the this group and for C_{12} we have $s_0 = h = 0$. So that the signature of Γ' is

$$(g', \pm, [2^{r_2}, 4^{r_4}, 8^{r_8}], \{(-)^{s_1}, (-)^{s_2}, (-)^{s_4}, (-)^{s_8})\}).$$

Then $k = s_1 + 2s_2 + 4s_4 + 8s_8$. By (1.3) we have

$$18 = 8\alpha g' + 4(r_2 + s_4) + 6(r_4 + s_2) + 7(r_8 + s_1). \tag{3.1}$$

As we made before, we are going to show six signatures which provide six of the eight classes modulo 8 of values of k:

a1	$(0, +, [2, 8^2], \{(-)^{s_8}\})$	$k = 8s_8$
a2	$(0, +, [2, 8], \{(-)^{s_1=1}, (-)^{s_8}\})$	$k = 8s_8 + 1$
a3	$(0, +, [2], \{(-)^{s_1=2}, (-)^{s_8}\})$	$k = 8s_8 + 2$
a4	$(0, +, [8^2], \{(-)^{s_4=1}, (-)^{s_8}\})$	$k = 8s_8 + 4$
a5	$(0, +, [8], \{(-)^{s_1=1}, (-)^{s_4=1}, (-)^{s_8}\})$	$k = 8s_8 + 5$
a6	$(0, +, [-], \{(-)^{s_1=2}, (-)^{s_4=1}, (-)^{s_8}\})$	$k = 8s_8 + 6$

The epimorphisms $\theta_1, \ldots, \theta_6$ are defined below. As in previous cases all relations are evidently satisfied.

$$\begin{aligned}
\theta_1 &: (x_1, x_2, x_3) \to (X^4, X, X^3) \\
\theta_2 &: (x_1, x_2, e_1) \to (X^4, X, X^3) \\
\theta_3 &: (x_1, e_1, e_2) \to (X^4, X, X^3) \\
\theta_4 &: (x_1, x_2, e_1) \to (X, X^3, X^4) \\
\theta_5 &: (x_1, e_1, e_2) \to (X, X^3, X^4) \\
\theta_6 &: (e_1, e_2, e_3) \to (X, X^3, X^4)
\end{aligned}$$

Let us observe that the classes 3 and 7 (mod 8) do not appear. For having such a k it would be necessary that $s_1 \geq 1$ and $s_2 \geq 1$, or else $s_1 \geq 3$, and therefore the above equality (3.1) cannot be fulfilled.

Group C_{10}

For this group first we take $s_0 = h = 0$ and so the signature of Γ' is

$$(g', \pm, [2^{r_2}, 5^{r_5}, 10^{r_{10}}], \{(-)^{s_1}, (-)^{s_2}, (-)^{s_5}, (-)^{s_{10}}\}).$$

Then $k = s_1 + 2s_2 + 5s_5 + 10s_{10}$. By (1.3) we have

$$22 = 10\alpha g' + 5(r_2 + s_5) + 8(r_5 + s_2) + 9(r_{10} + s_1).$$

Recall that the values of k congruent with 4 (mod 5) do not appear for the group C_5 and so they do not appear for the group C_{10}. Now we exhibit eight signatures which provide the eight classes (mod 10) excluding 4 and 9.

a1	$(0, +, [2, 5, 10], \{(-)^{s_{10}}\})$	$k = 10s_{10}$
a2	$(0, +, [2, 5], \{(-)^{s_1=1}, (-)^{s_{10}}\})$	$k = 10s_{10} + 1$
a3	$(0, +, [2, 10], \{(-)^{s_2=1}, (-)^{s_{10}}\})$	$k = 10s_{10} + 2$
a4	$(0, +, [2], \{(-)^{s_1=1}, (-)^{s_2=1}, (-)^{s_{10}}\})$	$k = 10s_{10} + 3$
a5	$(0, +, [5, 10], \{(-)^{s_5=1}, (-)^{s_{10}}\})$	$k = 10s_{10} + 5$
a6	$(0, +, [5], \{(-)^{s_1=1}, (-)^{s_5=1}, (-)^{s_{10}}\})$	$k = 10s_{10} + 6$
a7	$(0, +, [10], \{(-)^{s_2=1}, (-)^{s_5=1}, (-)^{s_{10}}\})$	$k = 10s_{10} + 7$
a8	$(0, +, [-], \{(-)^{s_1=1}, (-)^{s_2=1}, (-)^{s_5=1}, (-)^{s_{10}}\})$	$k = 10s_{10} + 8$

The epimorphisms $\theta_1, \ldots, \theta_8$ are defined below. As in previous cases all relations are evidently satisfied.

$$\theta_1 : (x_1, x_2, x_3) \to (X^5, X^4, X)$$
$$\theta_2 : (x_1, x_2, e_1) \to (X^5, X^4, X)$$
$$\theta_3 : (x_1, x_2, e_1) \to (X^5, X, X^4)$$
$$\theta_4 : (x_1, e_1, e_2) \to (X^5, X, X^4)$$
$$\theta_5 : (x_1, x_2, e_1) \to (X^4, X, X^5)$$
$$\theta_6 : (x_1, e_1, e_2) \to (X^4, X, X^5)$$
$$\theta_7 : (x_1, e_1, e_2) \to (X, X^4, X^5)$$
$$\theta_8 : (e_1, e_2, e_3) \to (X, X^4, X^5)$$

Group C_{12}

The signature of Γ' is

$$(g', \pm, (2^{r_2}, 3^{r_3}, 4^{r_4}, 6^{r_6}, 12^{r_{12}}), \{(-)^{s_1}, (-)^{s_2}, (-)^{s_3}, (-)^{s_4}, (-)^{s_6}, (-)^{s_{12}}\})$$

with $k = s_1 + 2s_2 + 3s_3 + 4s_4 + 6s_6 + 12s_{12}$, and applying (1.3) we have

$$26 = 12\alpha g' + 6(r_2 + s_6) + 8(r_3 + s_4) + 9(r_4 + s_3) + 10(r_6 + s_2) + 11(r_{12} + s_1)$$

which has the following solutions:

$\alpha = 2$	$g' = 0$	$r_{12} + s_1 = 1$	$r_4 + s_3 = 1$	$r_2 + s_6 = 1$
		$r_6 + s_2 = 2$	$r_2 + s_6 = 1$	
		$r_6 + s_2 = 1$	$r_3 + s_4 = 2$	
		$r_4 + s_3 = 2$	$r_3 + s_4 = 1$	
		$r_3 + s_4 = 1$	$r_2 + s_6 = 3$	
$\alpha = 1$	$g' = 1$	$r_3 + s_4 = 1$	$r_2 + s_6 = 1$	

The first five solutions do not allow to define the epimorphism θ. For instance, in the first case three elements of order 2, 4 and 12, respectively, do not have product equal to 1.

On the contrary, the last solution provides the following signatures:

a1	$(1, -, [2, 3], \{(-)^{s_{12}}\})$	$k = 12s_{12}$
a2	$(1, -, [2], \{(-)^{s_4=1}, (-)^{s_{12}}\})$	$k = 12s_{12} + 4$
a3	$(1, -, [3], \{(-)^{s_6=1}, (-)^{s_{12}}\})$	$k = 12s_{12} + 6$
a4	$(1, -, [-], \{(-)^{s_4=1}, (-)^{s_6=1}, (-)^{s_{12}}\})$	$k = 12s_{12} + 10$

for which the respective epimorphisms are defined by:

$$\theta_1 : (d_1, x_1, x_2) \to (X, X^6, X^4)$$
$$\theta_2 : (d_1, x_1, e_1) \to (X, X^6, X^4)$$
$$\theta_3 : (d_1, x_1, e_1) \to (X, X^4, X^6)$$
$$\theta_4 : (d_1, e_1, e_2) \to (X, X^4, X^6)$$

4. The other Abelian groups

For the groups $C_2 \times C_2$, $C_2 \times C_4$, $C_2 \times C_2 \times C_2$ and $C_2 \times C_6$ we deduce the possible signatures for Γ' from the Chapter 2 in [5]. If all period-cycles in the signature of Γ' are empty, from Theorem 2.3.2 in that book, we express their number as $s_{\tau_1} + s_{\tau_2} + \cdots + s_{\tau_\lambda} + s_0$, and then $k = \tau_1 s_{\tau_1} + \tau_2 s_{\tau_2} + \cdots + \tau_\lambda s_{\tau_\lambda}$, following the notation introduced in the beginning of Section 3. It is worth to note that in all groups appearing in the rest of the paper the equality (1.3) will always imply $s_0 = 0$. So, after $C_2 \times C_2$, we assume this value and delete this parameter from Γ'.

If there are non-empty period-cycles in the signature of Γ' the situation is a little more involved. The key point is that the non-empty period-cycles in Γ' can produce period-cycles of Γ. The number of the latter was obtained in Theorem 2.3.3 (see also the last sentence of Remark 2.3.7) in the book [5]. This result can be restated in the following way.

Theorem 4.1 *Let Γ be a bordered surface NEC group, N an even integer, and Γ' an NEC group containing Γ as a normal subgroup with quotient of order N. Let $(c_{i,0}, \ldots, c_{i,s_i})$ be one of the period-cycles of Γ', such that the reflection $c_{i,j} \in \Gamma$ while $c_{i,j-1}, c_{i,j+1} \notin \Gamma$ for $j \in J \subset \{1, 2, \ldots, s_i - 1\}$. Then the number of period-cycles of Γ produced by this period-cycle of Γ' is $\sum_{j\in J} \frac{N}{2n(j)}$, where we call $n(j)$ the order of $\Gamma(c_{i,j-1}c_{i,j+1})$.*

This result will be carefully described when dealing with the group $C_2 \times C_2$.

Group $C_2 \times C_2$

We call X and Y the generators of $C_2 \times C_2$. Firstly we consider that all period-cycles are empty. Then the signature of Γ' is

$$(g', \pm, [2^{r_2}], \{(-)^{s_2}, (-)^{s_4}, (-)^{s_0}\})$$

and then $k = 2s_2 + 4s_4$. Hence k can only be even. From (1.3) we have

$$10 = 4\alpha g' + 2(r_2 + s_2) + 4s_0.$$

We show two signatures which give all even values for k.

a1	$(0, +, [2^5], \{(-)^{s_4}\})$	$k = 4s_4$
a2	$(0, +, [2^4], \{(-)^{s_2=1}, (-)^{s_4}\})$	$k = 4s_4 + 2$

The respective epimorphisms θ_1 and θ_2 are defined by

$$\theta_1 : (x_1, x_2, x_3, x_4, x_5) \to (X, Y, XY, XY, XY)$$
$$\theta_2 : (x_1, x_2, x_3, x_4, e_1) \to (X, Y, XY, XY, XY)$$

In order to get the odd values of k it is necessary that there exist non-empty period-cycles in the signature of Γ'. Since Γ has only empty period-cycles there must appear two consecutive link-periods equal to 2 in a period-cycle of Γ' corresponding to two reflections which do not belong to the kernel of θ, see ([6]). In this case the signature of Γ' is

$$(g', \pm, [2^{r_2}], \{(-)^{s_2}, (-)^{s_4}, (-)^{s_0}, (2^{t_i})_{i=1,\dots,h}\})$$

with $h \geq 1$. Since we are looking just for odd values of k, let $k = 2s_2 + 4s_4 + 1$. Then applying (1.3) again we have

$$11 = 4\alpha g' + 2(r_2 + s_2) + 4(s_0 + h) + \sum_{i=1}^{h} t_i.$$

with $\sum_{i=1}^{h} t_i \geq 2$. A solution is $r_2 + s_2 = 2$, $h = 1$, $t_1 = 3$ which gives the two following signatures:

b1	$(0, +, [2^2], \{(2,2,2), (-)^{s_4}\})$	$k = 4s_4 + 1$
b2	$(0, +, [2], \{(2,2,2), (-)^{s_2=1}, (-)^{s_4}\})$	$k = 4s_4 + 3$

The respective epimorphisms are defined by

$$\theta_1 : (x_1, x_2, e_1, c_{1,0}, c_{1,1}, c_{1,2}, c_{1,3}) \to (X, X, 1, Y, 1, XY, Y)$$
$$\theta_2 : (x_1, e_1, e_2, c_{1,0}, c_{1,1}, c_{1,2}, c_{1,3}) \to (X, X, 1, Y, 1, XY, Y)$$

We can check easily that all relations are satisfied. However, it is important to note here that $\theta_j(c_{1,0}c_{1,2}) = X$ has order 2 and so this period-cycle produces $\frac{4}{2\cdot 2} = 1$ period-cycle in the signature of Γ. For this result see Theorem 4.1.

Group $C_2 \times C_4$

We call X and Y the generators of orders 2 and 4, respectively. First we suppose that all period-cycles of Γ' are empty, and so the signature is

$$(g', \pm, [2^{r_2}, 4^{r_4}], \{(-)^{s_2}, (-)^{s_4}, (-)^{s_8}\})$$

with $k = 2s_2 + 4s_4 + 8s_8$. From (1.3) we obtain

$$18 = 8\alpha g' + 4(r_2 + s_4) + 6(r_4 + s_2)$$

This provides three solutions as follows.

$\alpha = 2$ $\quad g' = 0$ $\quad r_4 + s_2 = 3$
$\alpha = 2$ $\quad g' = 0$ $\quad r_4 + s_2 = 1$ $\quad r_2 + s_4 = 3$
$\alpha = 1$ $\quad g' = 1$ $\quad r_2 + s_4 = 1$ $\quad r_4 + s_2 = 1$

However, the epimorphism θ cannot exist in any of these cases. Hence we need non-empty period-cycles in the signature of Γ'. Two elements of order 2 in $C_2 \times C_4$ have product of order 1 or 2. So we can get $\frac{8}{2\cdot 2} = 2$ or $\frac{8}{2\cdot 1} = 4$ empty period-cycles in Γ, and hence only even values of k can be obtained.

First we look for 2 period-cycles. Then $k = 2s_2 + 4s_4 + 8s_8 + 2$ and the signature of Γ' is

$$(g', \pm, [2^{r_2}, 4^{r_4}], \{(-)^{s_2}, (-)^{s_4}, (-)^{s_8}, (2^{t_i})_{i=1,\ldots,h}\})$$

Applying (1.3) we obtain

$$20 = 8\alpha g' + 4(r_2 + s_4) + 6(r_4 + s_2) + 8h + 2\sum_{i=1}^{h} t_i.$$

Since $h \geq 1$ and $\sum_{i=1}^{h} t_i \geq 2$, this implies $\alpha = 2$, $g' = 0$, $r_4 + s_2 = 1$, $h = 1$ and the period-cycle is $(2, 2, 2)$. So we have the signatures

a1	$(0, +, [4], \{(2, 2, 2), (-)^{s_8}\})$	$k = 8s_8 + 2$
a2	$(0, +, [-], \{(2, 2, 2), (-)^{s_2=1}, (-)^{s_8}\})$	$k = 8s_8 + 4$

The respective epimorphisms are defined by

$$\theta_1 : (x_1, e_1, c_{1,0}, c_{1,1,}, c_{1,2}, c_{1,3}) \to (Y, Y^3, X, 1, Y^2, X)$$
$$\theta_2 : (e_1, e_2, c_{1,0}, c_{1,1,}, c_{1,2}, c_{1,3}) \to (Y, Y^3, X, 1, Y^2, X)$$

Now we look for 4 period-cycles. We have

$$22 = 8\alpha g' + 4(r_2 + s_4) + 6(r_4 + s_2) + 8h + 2\sum_{i=1}^{h} t_i,$$

and then $\alpha = 2$, $g' = 0$, $h = 1$, $r_4 + s_2 = 1$, and the period-cycle is $(2, 2, 2, 2)$. We take the signature

b1 $(0, +, [-], \{(2, 2, 2, 2), (-)^{s_2=1}, (-)^{s_8}\})$ $k = 8s_8 + 6$

The epimorphism is defined by

$$\theta_1 : (e_1, e_2, c_{1,0}, c_{1,1,}, c_{1,2}, c_{1,3}, c_{1,4}) \to (Y, Y^3, X, 1, X, Y^2, X)$$

Finally we consider the case that the non-empty period-cycle in Γ' does not produce period-cycles in Γ. Then we have

$$18 = 8\alpha g' + 4(r_2 + s_4) + 6(r_4 + s_2) + 8h + 2\sum_{i=1}^{h} t_i.$$

and so we obtain $\alpha = 2$, $g' = 0$, $h = 1$, $r_4 + s_2 = 1$, and the period-cycle is $(2, 2)$. We take the signature

c1 $(0, +, [4], \{(2, 2), ((-)^{s_8}\})$ $k = 8s_8$

The epimorphism is defined by

$$\theta_1 : (x_1, e_1, c_{1,0}, c_{1,1,}, c_{1,2}) \to (Y, Y^3, X, Y^2, X)$$

Group $C_2 \times C_2 \times C_2$

We call X, Y and Z the three generators of G. We start once more with $h = 0$. Then the signature of Γ' is

$$(g', \pm, [2^{r_2}], \{(-)^{s_4}, (-)^{s_8}\})$$

with $k = 4s_4 + 8s_8$. Hence applying (1.3) we obtain

$$18 = 8\alpha g' + 4(r_2 + s_4)$$

what has no solution. So we must take non-empty period-cycles in Γ' which provide 2 or 4 empty period-cycles in Γ. The signature of Γ' is now

$$(g', \pm, [2^{r_2}], \{(-)^{s_4}, (-)^{s_8}, (2^{t_i})_{i=1,\dots,h}\}).$$

First we look for 2 period-cycles. From (1.3) we have

$$20 = 8\alpha g' + 4(r_2 + s_4) + 8h + 2\sum_{i=1}^{h} t_i$$

We choose two signatures satisfying this equality, namely

a1	$(0, +, [2], \{(2^4), (-)^{s_8}\})$	$k = 8s_8 + 2$
a2	$(0, +, [-], \{(2^4), (-)^{s_4=1}, (-)^{s_8}\})$	$k = 8s_8 + 6$

The respective epimorphisms are defined by

$$\theta_1 : (x_1, e_1, c_{1,0}, c_{1,1}, c_{1,2}, c_{1,3}, c_{1,4}) \to (X, X, Y, 1, Z, XZ, Y)$$
$$\theta_2 : (e_1, e_2, c_{1,0}, c_{1,1}, c_{1,2}, c_{1,3}, c_{1,4}) \to (X, X, Y, 1, Z, XZ, Y)$$

Now we look for 4 period-cycles. Then we have

$$22 = 8\alpha g' + 4(r_2 + s_4) + 8h + 2\sum_{i=1}^{h} t_i$$

We select the following two signatures:

b1	$(0, +, [2], \{(2^5), (-)^{s_8}\})$	$k = 8s_8 + 4$
b2	$(0, +, [-], \{(2^5), (-)^{s_4=1}, (-)^{s_8}\})$	$k = 8s_8 + 8$

and the epimorphisms are defined by

$$\theta_1 : (x_1, e_1, c_{1,0}, c_{1,1}, c_{1,2}, c_{1,3}, c_{1,4}, c_{1,5}) \to (X, X, Y, 1, Y, Z, XZ, Y)$$
$$\theta_2 : (e_1, e_2, c_{1,0}, c_{1,1}, c_{1,2}, c_{1,3}, c_{1,4}, c_{1,5}) \to (X, X, Y, 1, Y, Z, XZ, Y)$$

So the group $C_2 \times C_2 \times C_2$ acts on surfaces with all even values of k.

Group $C_2 \times C_6$

We call X and Y the generators of orders 2 and 6 respectively. For simplicity we take $s_0 = 0$ and firstly we consider that all period-cycles are empty. Then the signature of Γ' is

$$(g', \pm, [2^{r_2}, 3^{r_3}, 6^{r_6}], \{(-)^{s_2}, (-)^{s_4}, (-)^{s_6}, (-)^{s_{12}}\})$$

and then $k = 2s_2 + 4s_4 + 6s_6 + 12s_{12}$ and so it is even. From (1.3) we have

$$26 = 12\alpha g' + 6(r_2 + s_6) + 8(r_3 + s_4) + 10(r_6 + s_2).$$

We show now six signatures satisfying this equality which give all even values of k.

a1	$(0, +, [2, 6^2], \{(-)^{s_{12}}\})$	$k = 12s_{12}$
a2	$(0, +, [2, 6], \{(-)^{s_2=1}, (-)^{s_{12}}\})$	$k = 12s_{12} + 2$
a3	$(0, +, [2], \{(-)^{s_2=2}, (-)^{s_{12}}\})$	$k = 12s_{12} + 4$
a4	$(0, +, [6^2], \{(-)^{s_6=1}, (-)^{s_{12}}\})$	$k = 12s_{12} + 6$
a5	$(0, +, [6], \{(-)^{s_2=1}, (-)^{s_6=1}, (-)^{s_{12}}\})$	$k = 12s_{12} + 8$
a6	$(0, +, [-], \{(-)^{s_2=2}, (-)^{s_6=1}, (-)^{s_{12}}\})$	$k = 12s_{12} + 10$

The respective epimorphisms $\theta_1, \ldots, \theta_6$ are

$$\begin{aligned}
\theta_1 &: (x_1, x_2, x_3) \to (X, Y, XY^5)\\
\theta_2 &: (x_1, x_2, e_1) \to (X, Y, XY^5)\\
\theta_3 &: (x_1, e_1, e_2) \to (X, Y, XY^5)\\
\theta_4 &: (x_1, x_2, e_1) \to (Y, XY^5, X)\\
\theta_5 &: (x_1, e_1, e_2) \to (Y, XY^5, X)\\
\theta_6 &: (e_1, e_2, e_r) \to (Y, XY^5, X)
\end{aligned}$$

Now we look for the odd values of k. Two elements of order 2 in the group $C_2 \times C_6$ cannot have a product of order 6. Therefore a non-empty period-cycle in Γ' does not produce $\frac{12}{2 \cdot 6} = 1$ empty period-cycle in Γ. Hence we are going to take $k = 2s_2 + 4s_4 + 6s_6 + 12s_{12} + 3$. The signature of Γ' is

$$(g', \pm, [2^{r_2}, 3^{r_3}, 6^{r_6}], \{(-)^{s_2}, (-)^{s_4}, (-)^{s_6}, (-)^{s_{12}}, (2^{t_i})_{i=1,\ldots,h}\}).$$

Then applying (1.3) we have

$$29 = 12\alpha g' + 6(r_2 + s_6) + 8(r_3 + s_4) + 10(r_6 + s_2) + 12h + 3\sum_{i=1}^{h} t_i.$$

Since $h \geq 1$, and $\sum_{i=1}^{h} t_i \geq 2$, the unique solution comes from $r_3 + s_4 = 1$, $h = 1$, $t_1 = 3$.

Then we have two signatures:

b1	$(0,+,[3],\{(2,2,2),(-)^{s_{12}}\})$	$k=12s_{12}+3$
b2	$(0,+,[-],\{(2,2,2),(-)^{s_4=1},(-)^{s_{12}}\})$	$k=12s_{12}+7$

The respective epimorphisms are defined by

$$\theta_1:(x_1,e_1,c_{1,0},c_{1,1},c_{1,2},c_{1,3})\to(Y^2,Y^4,X,1,Y^3,X)$$
$$\theta_2:(e_1,e_2,c_{1,0},c_{1,1},c_{1,2},c_{1,3})\to(Y^2,Y^4,X,1,Y^3,X)$$

We have proven that the only odd values of k are the numbers congruent to 3 or 7 (mod 12). So that, we have finished the study of possible Abelian groups.

5. The global result

We collect all results obtained for the different groups in the following

Theorem 5.1 *Let G be an Abelian group of automorphisms of an orientable Klein surface of topological genus 2 with $k>0$ boundary components. Then the group G and the admissible values of k for each G are given in the next Table.*

g	k
C_2	All k
C_3	All k
C_4	All k
$C_2\times C_2$	All k
C_5	$k\equiv 0,1,2,3 \pmod 5$
C_6	All k
C_8	$k\equiv 0,1,2 \pmod 4$
$C_2\times C_4$	k even
$C_2\times C_2\times C_2$	k even
C_{10}	$k\equiv 0,1,2,3 \pmod 5$
C_{12}	$k\equiv 0,4,6,10 \pmod{12}$
$C_2\times C_6$	$k\equiv 0,2,3,4,6,7,8,10 \pmod{12}$

Remark 5.2 *Let us observe that the automorphism groups of Klein surfaces with one boundary component were obtained in Chapter 5 of [5]. The result in the above Theorem matches with theorem 5.2.3 of that book.*

References

[1] N. L. Alling, N. Greenleaf: *Foundations of the theory of Klein surfaces.* LNM **219,** Springer-Verlag 1971.

[2] S. A. Broughton: *Classifying finite group actions on surfaces of low genus*. J. Pure Appl. Algebra **69** (1990) 233–270.

[3] E. Bujalance: *Automorphism groups of compact planar Klein surfaces*. Manuscripta Math. **56** (1986) 105–124.

[4] E. Bujalance, M. D. Conder, A. F. Costa: *Pseudo-real Riemann surfaces and chiral regular maps*. Trans. AMS **362** (2010) 3365–3376.

[5] E. Bujalance, J. J. Etayo, J. M. Gamboa, G. Gromadzki: *Automorphism Groups of Compact Bordered Klein Surfaces. A Combinatorial Approach*. LNM **1439,** Springer-Verlag 1990.

[6] E. Bujalance, E. Martínez: *A remark on NEC groups representing surfaces with boundary*. Bull. London Math. Soc. **21** (1989) 263–266.

[7] E. Bujalance, D. Singerman: *The symmetry type of a Riemann surface*. Proc. London Math. Soc. (3) **51** (1985) 501–519.

[8] J. A. Bujalance: *Normal subgroups of even index in an NEC group*. Archiv der Math. **49** (1987) 470–478.

[9] N. Greenleaf, C. L. May: *Bordered Klein surfaces with maximal symmetry*. Trans. AMS **274** (1982) 265–283.

[10] A. M. Macbeath: *The classification of non-Euclidean crystallographic groups*. Canad. J. Math. **19** (1967) 1192–1205.

[11] C. L. May: *Large automorphism groups of compact Klein surfaces with boundary*. Glasgow Math. J. **18** (1977) 1–10.

[12] R. Preston: *Projective structures and foundamental domains on compact Klein surfaces*. Doctoral Thesis. University of Texas, 1975.

[13] H. C. Wilkie: *On non-Euclidean crystallographic groups*. Math. Z. **91** (1966) 87–102.

Classifying PL 4-manifolds via crystallizations: results and open problems

Maria Rita Casali, Paola Cristofori and Carlo Gagliardi*

Dipartimento di Scienze Fisiche, Informatiche e Matematiche
Università di Modena e Reggio Emilia
Via Campi 213 B
I-41125 MODENA (Italy)

mariarita.casali@unimore.it paola.cristofori@unimore.it carlo.gagliardi@unimore.it

To our friend José María, with admiration and gratitude.

Abstract

Crystallization theory is a graph-theoretical representation method for compact PL-manifolds of arbitrary dimension, which makes use of a particular class of edge-coloured graphs, which are dual to coloured (pseudo-)triangulations. The purely combinatorial nature of crystallizations makes them particularly suitable for automatic generation and classification, as well as for the introduction and study of graph-defined invariants for PL-manifolds.

The present survey paper focuses on the 4-dimensional case, presenting up-to-date results about the PL classification of closed 4-manifolds, by means of two such PL invariants: *regular genus* and *gem-complexity.*

Open problems are also presented, mainly concerning different classification of 4-manifolds in TOP and DIFF=PL categories, and a possible approach to the 4-dimensional Smooth Poincaré Conjecture.

2010 Mathematics Subject Classification: 57Q15 - 57N13 - 57M15.
Key words: PL 4-manifold, coloured graph, coloured triangulation, regular genus, gem-complexity, simple crystallization, semi-simple crystallization.

*Work supported by the "National Group for Algebraic and Geometric Structures, and their Applications" (GNSAGA - INDAM) and by M.I.U.R. of Italy (project "Strutture Geometriche, Combinatoria e loro Applicazioni").

1. Introduction

Crystallization theory is a representation theory for PL-manifolds by means of edge-coloured graphs, which are dual 1-skeletons of particular (pseudo)-triangulations. These graphs are called *crystallizations* if the associated triangulations have the minimum possible number of vertices.

The peculiarity of this method consists in its universality: in fact, it enables to represent and study the whole class of PL-manifolds, without restrictions about dimension, boundary property and orientability (despite what happens with other classical representation theories, whose extension beyond dimension three appears to be quite difficult to achieve, or restricted to particular hypotheses).

The possibility of performing a purely combinatorial approach to general PL-manifolds is of particular interest in dimension greater or equal to four, where the difference between the categories TOP and PL (and DIFF, if $n \geq 5$) must be taken into account. For example, it is very useful to have combinatorial moves on the representing objects, which realize the PL-homeomorphism (and not only the TOP-homeomorphism) between the represented manifolds, or to define PL invariants of the manifolds (possibly distinguishing different PL structures on the same TOP-manifold), whose computation can be performed directly on the combinatorial objects.

Within crystallization theory, both tools are available: suitable sets of moves on edge-coloured graphs exist, which enable to recognize different crystallizations of the same PL n-manifold, and some interesting graph-defined invariants for PL-manifolds have been introduced and deeply studied.

In particular:

- the *gem-complexity* $k(M^n)$ of a PL n-manifold M^n is the integer $p-1$, where $2p$ is the minimum order of a crystallization of M^n;

- the *regular genus* $\mathcal{G}(M^n)$ of an orientable (resp. non-orientable) PL n-manifold M^n is defined as the minimum genus (resp. half the minimum genus) of a surface into which a crystallization of M^n regularly embeds (see Section 2).

Note that the regular genus extends to higher dimension the classical concept of Heegaard genus of a 3-manifold, and succeeds in characterizing PL-spheres and disks of arbitrary dimension (see Subsection 4.1); on the other hand, gem-complexity is the natural invariant involved in possible generation and analysis of catalogues of PL n-manifolds via crystallizations.[1]

The present survey paper focuses on the 4-dimensional closed case, and presents in a unified view updated results - most of them very recent and still in publication -

[1]In dimension three, the automatic generation and analysis of catalogues of 3-dimensional crystallizations for increasing values of their vertex-number has already produced the classification of all closed 3-manifolds up to gem-complexity 14: see [4], [18].

about topological and PL classifications of closed PL 4-manifolds according to both gem-complexity and regular genus. Some new results are also included.

As regards regular genus, the classifying results related to closed PL 4-manifolds mainly concern the case of "low" regular genus (Proposition 4.2) and the case of "restricted gap" between the regular genus and the rank of the fundamental group of the manifold (Proposition 4.3). All of them are based essentially on the existence of particular types of handle decompositions induced by crystallizations of the involved PL 4-manifolds, and make use of an important result by Montesinos [45] ensuring the uniqueness of the boundary identification of two handlebodies (see Subsection 4.1).

On the other hand, the classifying results via gem-complexity are based on the development of an effective algorithm for the automatic generation and classification of the catalogue of 4-manifold crystallizations up to a fixed number of vertices (see Subsection 4.2). Theorem 4.6 summarizes the complete PL classification of closed orientable (resp. non-orientable) PL 4-manifolds up to gem-complexity 8 (resp. up to gem-complexity 9), due to [20], and contains also a new statement, concerning the partial - and in progress - classification of crystallizations with 20 vertices.

The difficulty of the exact calculation of both the regular genus and gem-complexity for any given PL 4-manifold, makes the search for significant lower and upper bounds a relevant task. In Section 3 a result - recently obtained in [6] - is presented, yielding sharp lower bounds for both invariants, by means of the Euler characteristic and the rank of the fundamental group of the involved 4-manifold (Theorem 3.1). These bounds turn out to be very useful to improve estimation for regular genus and gem-complexity of product 4-manifolds (see Subsection 3.2), and to obtain a new proof of the TOP classification of simply-connected 4-manifolds up to regular genus 43 and gem-complexity 65 (Theorem 3.5).

Section 5 is devoted to the so called *semi-simple crystallizations*, introduced in [6] so that the represented PL 4-manifolds attain the above lower bounds. The additivity of both gem-complexity and regular genus with respect to connected sum is proved for such a class of PL 4-manifolds, which comprehends all "standard" ones and their connected sums.

Note that additivity of regular genus for closed PL 4-manifolds was conjectured in [31] and has been proved to imply - by a theorem of Wall - the 4-dimensional Smooth Poincaré Conjecture. Therefore, the identification of classes of manifolds for which the property holds is an interesting open problem (see Section 6.2).

Other further developments, mainly concerning different classification of 4-manifolds in TOP and DIFF=PL categories, are reviewed in the last section of the paper. In particular, it is discussed the possible application of the classification algorithm to the crystallizations arising from the two known 16-vertices and 17-vertices triangulations of the $K3$-surface obtained in [26] and [46] respectively.

2. Basic notions of crystallization theory

In the present work, when not otherwise explicitly specified, we will consider only closed, connected piecewise linear manifolds of dimension $n = 4$ (simply referred to as "PL 4-manifolds"). Therefore, although edge-coloured graphs are a representation tool for the whole class of PL-manifolds, in this section we will briefly review basic notions and results of the theory with respect to this particular case.

A *5-coloured graph (without boundary)* is a pair (Γ, γ), where $\Gamma = (V(\Gamma), E(\Gamma))$ is a regular multigraph (i.e. it may include multiple edges, but no loop) of degree five and $\gamma : E(\Gamma) \to \Delta_4 = \{0, 1, 2, 3, 4\}$ is a proper edge-coloration (i.e. it is injective when restricted to the set of edges incident to any vertex of Γ).

The elements of the set Δ_4 are called the *colours* of Γ; thus, for every $i \in \Delta_4$, an *i-coloured edge* is an element $e \in E(\Gamma)$ such that $\gamma(e) = i$. For every $i, j, k \in \Delta_4$ let $\Gamma_{\hat{\imath}}$ (resp. Γ_{ijk}) (resp. Γ_{ij}) be the subgraph obtained from (Γ, γ) by deleting all the edges of colour i (resp. $c \in \Delta_4 - \{i, j, k\}$) (resp. $c \in \Delta_4 - \{i, j\}$). The connected components of $\Gamma_{\hat{\imath}}$ (resp. Γ_{ijk}) (resp. Γ_{ij}) are called *$\hat{\imath}$-residues* (resp. *$\{i, j, k\}$-coloured residues*) (resp. *$\{i, j\}$-coloured cycles*) of Γ, and their number is denoted by $g_{\hat{\imath}}$ (resp. g_{ijk}) (resp. g_{ij}).

A 5-coloured graph (Γ, γ) is called *contracted* iff, for each $i \in \Delta_4$, the subgraph $\Gamma_{\hat{\imath}}$ is connected (i.e. iff $g_{\hat{\imath}} = 1 \ \forall i \in \Delta_4$).

Every 5-coloured graph (Γ, γ) may be thought of as the combinatorial visualization of a 4-dimensional labelled pseudocomplex $K(\Gamma)$, which is constructed in the following way:

- for each vertex $v \in V(\Gamma)$, take a 4-simplex $\sigma(v)$, with vertices labelled $0, 1, 2, 3, 4$;
- for each j-coloured edge between v and w ($v, w \in V(\Gamma)$), identify the 3-dimensional faces of $\sigma(v)$ and $\sigma(w)$ opposite to the vertex labelled j, so that equally labelled vertices coincide.

In case $K(\Gamma)$ triangulates a PL 4-manifold M, then (Γ, γ) is called a *gem* (gem = graph encoded manifold) *representing* M.

In the following, for sake of conciseness, we will write Γ instead of (Γ, γ), when there is no ambiguity with regard to the edge-coloration.

The following proposition summarizes some useful results which come directly from the above construction.

Proposition 2.1 *If Γ is an order $2p$ gem of a PL 4-manifold M, then:*

(a) M is orientable iff Γ is bipartite;

(b) there is a bijection between i-labelled vertices (resp. 1-simplices whose vertices are labelled $\Delta_4 - \{i, j, k\}$) (resp. 2-simplices whose vertices are labelled $\Delta_4 - \{i, j\}$) of $K(\Gamma)$ and $\hat{\imath}$-residues (resp. $\{i, j, k\}$-coloured residues) (resp. $\{i, j\}$-coloured cycles) of Γ;

(c) $\chi(|K(\Gamma)|) = -3p + \sum_{i,j} g_{ij} - \sum_{i,j,k} g_{ijk} + \sum_i g_i$;

(d) $2g_{ijk} = g_{ij} + g_{ik} + g_{jk} - p$ *for each triple* $(i,j,k) \in \Delta_4$;

(e) *for each distinct* $i,j,k \in \Delta_4$, *there exists a presentation of* $\pi_1(M)$ *whose generators are in bijection with the connected components of* Γ_{ijk} *but one.*

A gem representing a PL 4-manifold M is a *crystallization* of M if it is also a contracted graph; by the above property (b), this is equivalent to require that the associated pseudocomplex $K(\Gamma)$ contains exactly five vertices (one for each label $i \in \Delta_4$). Pezzana Theorem and its subsequent improvements prove that every PL-manifold admits a crystallization (see [32]).

The following proposition allows to characterize crystallizations of PL 4-manifolds among 5-coloured graphs.

Proposition 2.2 *A 5-coloured graph* Γ *is a crystallization of a PL 4-manifold if and only if, for every* $c \in \Delta_4$, $\Gamma_{\hat{c}}$ *is connected and represents* $\mathbb{S}^3$.

Catalogues of crystallizations of PL manifolds have been obtained both in dimension three (see [41], [17] and [18] for the 3-dimensional orientable case and [14], [16] and [4] for the non-orientable one) and four [20]. As mentioned in Section 1, they are constructed with respect to a suitable graph-defined PL invariant, which measures how "complicated" the representing combinatorial object is[2].

Definition 1 Given a PL n-manifold M^n, its *gem-complexity* is the non-negative integer $k(M^n) = p - 1$, where $2p$ is the minimum order of a crystallization of M^n.

An *h-dipole* ($1 \leq h \leq 4$) of a 5-coloured graph Γ is a subgraph of order two of Γ, having h edges coloured by $\{c_1, \dots, c_h\}$, such that its vertices belong to different connected components of $\Gamma_{\Delta_4 - \{c_1,\dots,c_h\}}$.

A *ρ-pair* in Γ is a pair of equally coloured edges both belonging to at least three common bicoloured cycles of Γ.

It is proved in [20, Proposition 20] that, if M is a *handle-free* PL 4-manifold (i.e.: if it admits neither the orientable nor the non-orientable $\mathbb{S}^3$-bundle over $\mathbb{S}^1$ as a connected summand), then $k(M) = p - 1$, where $2p$ is the order of a crystallization of M with no dipoles and no ρ-pairs.

Crystallizations with these properties are called *rigid dipole-free crystallizations*; they are exactly the elements considered in the existing crystallization catalogues in dimension four.[3]

Another graph-based invariant for PL n-manifolds, called *regular genus*, is related to some of the most interesting results of crystallization theory[4]. It was introduced

[2] In dimension three, the relations between this invariant and the well-known Matveev's complexity have been widely investigated: see [16], [17], [19] and [22].

[3] A slightly modified definition of *rigidity* is required in 3-dimensional crystallization catalogues.

[4] See, for example, [11], [12] and [25] for 4-dimensional results, [23] and [13] for 5-dimensional ones.

in [35] and its definition relies on the existence of a particular type of embedding into a surface for gems of arbitrary dimension.

As far as the 4-dimensional case is concerned, it is well-known that, if Γ is an order $2p$ crystallization of an orientable (resp. non-orientable) PL 4-manifold M, then for every cyclic permutation $\varepsilon = (\varepsilon_o, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4 = 4)$ of Δ_4 there exists a so-called *regular embedding*[5] $i_\varepsilon : |\Gamma| \to F_\varepsilon$, where F_ε is the closed orientable (resp. non-orientable) surface of genus $\rho_\varepsilon(\Gamma)$ (resp. $2\rho_\varepsilon(\Gamma)$), where $\rho_\varepsilon(\Gamma)$ may be directly computed by the following formula (see [35] for details and suitable extensions in general dimension).

$$\sum_{i\in\mathbb{Z}_5} g_{\varepsilon_i\varepsilon_{i+1}} - 3p = 2 - 2\rho_\varepsilon(\Gamma). \tag{2.1}$$

Definition 2 The *regular genus* of a bipartite (resp. non-bipartite) 5-coloured graph Γ is defined as the minimum genus (resp. half the minimum genus) of a surface into which Γ regularly embeds:

$$\rho(\Gamma) = \min_{\varepsilon}\{\rho_\varepsilon(\Gamma)\};$$

the *regular genus* of a PL 4-manifold M is defined as the minimum regular genus of a crystallization of M:

$$\mathcal{G}(M) = \min\{\rho(\Gamma) \ / \ (\Gamma,\gamma) \text{ crystallization of } M\}.$$

Finally, we recall that, given two 5-coloured graphs Γ_1 and Γ_2 representing PL 4-manifolds M_1 and M_2 respectively, for any choice of $v_1 \in V(\Gamma_1)$ and $v_2 \in V(\Gamma_2)$, it is possible to construct a new 5-coloured graph $\Gamma_1\#_{v_1,v_2}\Gamma_2$, called a *graph connected sum* of Γ_1 and Γ_2, by deleting v_1 and v_2 and welding the hanging edges according to their colours.

$\Gamma_1\#_{v_1,v_2}\Gamma_2$ turns out to be a gem of one of the two possible connected sums of M_1 and M_2 (see [32] for details).

3. Lower bounds and their consequences

3.1. Lower bounds for regular genus and gem-complexity

The following result - recently obtained by Basak and Casali - is very useful to investigate PL-manifolds of dimension four by means of the two invariants regular genus and gem-complexity.

Theorem 3.1 [6] *Let M be a PL 4-manifold with $rk(\pi_1(M)) = m$. Then:*

$$k(M) \ \geq \ 3\chi(M) + 10m - 6;$$

$$\mathcal{G}(M) \ \geq \ 2\chi(M) + 5m - 4.$$

[5] By short, it is a cellular embedding whose regions are bounded by the images of $\{\varepsilon_i, \varepsilon_{i+1}\}$-coloured cycles, for each $i \in \mathbb{Z}_5$.

Roughly speaking, we can say that the proof of the inequality concerning gem-complexity is based on the Dehn-Sommerville equations in dimension four, applied to the contracted pseudo-triangulation $K(\Gamma)$ of M associated to any crystallization of M, by making use of the relations $g_{ijk} \geq m+1$ for any distinct $i, j, k \in \Delta_4$ (coming from the assumption $rk(\pi_1(M)) = m$ and Proposition 2.1(e)).

On the contrary, the inequality concerning regular genus makes use of the following crucial steps (see [6, Theorem 1] for details):

- By the first inequality of Theorem 3.1, the minimum possible order of a crystallization of M is $2\bar{p} = 6\chi(M) + 10(2m-1)$; hence, any crystallization (Γ, γ) of M has $\#V(\Gamma) = 2\bar{p} + 2q$ for some non-negative integer q.
- For any distinct $i, j, k \in \Delta_4$, the assumption $rk(\pi_1(M)) = m$ implies $g_{ijk} = (m+1) + t_{ijk}$, where $t_{ijk} \in \mathbb{Z}$, $t_{ijk} \geq 0$ and $\sum_{0 \leq i<j<k \leq 4} t_{ijk} = q$.
- By Proposition 2.1(d), the ten relations $g_{ij} + g_{ik} + g_{jk} = 2g_{ijk} + (\bar{p} + q)$ $(0 \leq i < j < k \leq 4)$ give rise to a linear system of equations (in the numbers of the different bicoloured cycles) which may be solved, so to obtain the following lower bound (surprisingly not depending from q) for the regular genus of Γ with respect to any cyclic permutation ε of Δ_4: $\rho_\varepsilon(\Gamma) \geq \frac{2(\bar{p}-1)-5m}{3}$.
- Since both the crystallization Γ of M and the permutation ε of Δ_4 are arbitrary, the second inequality of Theorem 3.1 easily follows.

3.2. Regular genus and gem-complexity of product 4-manifolds

In [6], Theorem 3.1 is applied in order to significantly improve some lower bounds for the regular genus of PL 4-manifolds, which have been proved by various authors via different techniques; meanwhile, similar lower bounds are obtained also for gem-complexity.

Proposition 3.2 [6] *For any closed 3-manifold M^3 such that $\pi_1(M^3)$ is a finite abelian group, then:*

$$\mathcal{G}(M^3 \times \mathbb{S}^1) \geq 5rk(\pi_1(M^3)) + 1 \qquad \textit{and} \qquad k(M^3 \times \mathbb{S}^1) \geq 10rk(\pi_1(M^3)) - 6.$$

In particular:

$$\mathcal{G}(L(p,q) \times \mathbb{S}^1) \geq 6 \qquad \textit{and} \qquad k(L(p,q) \times \mathbb{S}^1) \geq 4.$$

Proposition 3.3 [6] *Let T_g (resp. U_h) denote the orientable (resp. non-orientable) surface of genus $g \geq 0$ (resp. $h \geq 1$). Then:*

$$\mathcal{G}(T_g \times T_r) \geq 8gr + 2g + 2r + 4 \qquad \textit{and} \qquad k(T_g \times T_r) \geq 12gr + 8g + 8r + 6;$$

$$\mathcal{G}(T_g \times U_h) \geq 4gh + 2g + h + 4 \qquad \textit{and} \qquad k(T_g \times U_h) \geq 6gh + 8g + 4h + 6;$$

$$\mathcal{G}(U_h \times U_k) \geq 2hk + h + k + 4 \qquad and \qquad k(U_h \times U_k) \geq 3hk + 4h + 4k + 6.$$

In particular:

$$\mathcal{G}(\mathbb{S}^2 \times T_g) \geq 2g + 4 \qquad and \qquad k(\mathbb{S}^2 \times T_g) \geq 8g + 6;$$

$$\mathcal{G}(\mathbb{S}^2 \times U_h) \geq h + 4 \qquad and \qquad k(\mathbb{S}^2 \times U_h) \geq 4h + 6.$$

Moreover, the last inequality of Proposition 3.3 concerning regular genus (resp. gem-complexity), together with the existence of a genus five (resp. order 24) crystallization of $\mathbb{S}^2 \times \mathbb{RP}^2$ depicted in [6, Figure 3], allows the exact calculation of the regular genus (resp. an estimation with “strict range” of the gem-complexity) of the involved 4-manifold:

Proposition 3.4 [6]

$$\mathcal{G}(\mathbb{S}^2 \times \mathbb{RP}^2) = 5 \qquad and \qquad k(\mathbb{S}^2 \times \mathbb{RP}^2) \in \{10, 11\}.$$

3.3. TOP classification of simply-connected 4-manifolds via regular genus and gem-complexity

A direct application of Theorem 3.1 to the case of simply-connected PL 4-manifolds, combined with well-known results on TOP simply-connected 4-manifolds, yields the following interesting result related to the topological classification of simply-connected PL 4-manifolds with respect both to *gem-complexity* and to *regular genus*:[6]

Theorem 3.5 [20] *Let M be a simply-connected PL 4-manifold. If either $k(M) \leq 65$ or $\mathcal{G}(M) \leq 43$, then M is TOP-homeomorphic to*

$$(\#_r \mathbb{CP}^2)\#(\#_{r'}(-\mathbb{CP}^2)) \quad or \quad \#_s(\mathbb{S}^2 \times \mathbb{S}^2),$$

where $r + r' = \beta_2(M)$, $s = \frac{1}{2}\beta_2(M)$ and $\beta_2(M)$ is the second Betti number of M.

Proof. Since M is assumed to be simply-connected, Theorem 3.1 yields:

$$k(M) \geq 3\beta_2(M) \tag{3.1}$$

and

$$\mathcal{G}(M) \geq 2\beta_2(M). \tag{3.2}$$

[6]Note that the proof presented in the present survey paper is easier than the original one (contained in [20, Proposition 20 and Proposition 23]), since it directly makes use of the inequalities derived from Theorem 3.1.

Now, the classical theorems of Freedman and Donaldson (see [33]) about the TOP classification of simply-connected closed 4-manifolds, together with more recent results by Furuta [34], ensure that intersection forms of type

$$\pm 2nE_8 \oplus s \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

do represent a PL 4-manifold only if $s > 2n$; hence, only PL 4-manifolds with $\beta_2(M) \geq 22$ occur in this case. The thesis directly follows from the fact that both $k(M) \leq 65$ and $\mathcal{G}(M) \leq 43$ imply $\beta_2(M) \leq 21$; so, only intersection forms of the two simplest types are allowed:

$$r[1] \oplus r'[-1] \qquad or \qquad s \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

where $r + r' = \beta_2(M)$ or $s = \frac{1}{2}\beta_2(M)$. □

4. PL classification via regular genus and gem-complexity

4.1. Classifying results via regular genus

The first, important property of regular genus consists in the possibility of recognizing spheres (and disks[7]) of arbitrary dimension, in full analogy with well-known low-dimensional characterizations.

Theorem 4.1 [31] *For every closed PL n-manifold M^n, with $n \geq 2$,*

$$\mathcal{G}(M^n) = 0 \iff M^n \cong \mathbb{S}^n$$

Actually, the regular genus shares many properties with other low-dimensional genera (see [3] for a survey on results in general dimension, including the boundary case); for example, for every PL n-manifold M^n, $n \geq 3$, $\mathcal{G}(M^n)$ is a non-negative integer invariant, so that $\mathcal{G}(M^n) \geq rk(\pi_1(M^n))$.

Starting from the above results, many efforts have been spent in order to investigate the relation existing between the "PL structure" of a manifold M^n and its regular genus $\mathcal{G}(M^n)$, in order to yield classifying results via regular genus in PL category and dimension n, both in the closed and in the boundary case. The best results have been achieved in dimension 4 and 5, and concern the case of "low" regular genus, the case of "restricted gap" between the regular genus of the manifold and the regular genus of its boundary, and the case of "restricted gap" between the regular genus and the rank of the fundamental group of the manifold.

[7]Several results of this subsection admit suitable extensions to PL manifolds with non-empty boundary. However, for sake of conciseness, we restrict the statement of Theorem 4.1 to the closed case, as everywhere in the paper.

The following Propositions 4.2 and 4.3 exactly deal with the first and third cases in the closed 4-dimensional setting, while the second one, which concerns PL 4-manifolds with boundary, is out of the scope of the present survey.

From now on, $\mathbb{S}^1 \times \mathbb{S}^3$ (resp. $\mathbb{S}^1 \tilde{\times} \mathbb{S}^3$) (resp. $\mathbb{S}^1 \otimes \mathbb{S}^3$) will denote the orientable (resp. non-orientable) (resp. either orientable or non-orientable) $\mathbb{S}^3$-bundle over $\mathbb{S}^1$.

Proposition 4.2 [27, 28, 29]

a) Let M be a closed orientable 4-manifold; then:

$$\mathcal{G}(M) = \rho \leq 3 \quad \Longrightarrow \quad M \cong \begin{cases} \#_\rho(\mathbb{S}^1 \times \mathbb{S}^3) \\ \#_{\rho-2}(\mathbb{S}^1 \times \mathbb{S}^3)\#\mathbb{CP}^2 \end{cases}$$

b) Let M be a closed non-orientable 4-manifold; then:

$$\mathcal{G}(M) = \rho \leq 2 \quad \Longrightarrow \quad M \cong \#_\rho(\mathbb{S}^1 \tilde{\times} \mathbb{S}^3)$$

Proposition 4.3 [12, 25]

a) Let M be a closed (orientable or non-orientable) 4-manifold; then:

$$\mathcal{G}(M) = rk(\pi_1(M)) = \rho \quad \Longleftrightarrow \quad M \cong \#_\rho(\mathbb{S}^1 \otimes \mathbb{S}^3)$$

b) Let M be a closed 4-manifold; then:

- $\mathcal{G}(M) \neq rk(\pi_1(M)) \implies \mathcal{G}(M) - rk(\pi_1(M)) \geq 2$
- $\mathcal{G}(M) - rk(\pi_1(M)) = 2$ *and* $\pi_1(M) = *_m\mathbb{Z} \iff M \cong \#_m(\mathbb{S}^1 \otimes \mathbb{S}^3)\#\mathbb{CP}^2$
- *No closed 4-manifold M exists with* $\mathcal{G}(M) - rk(\pi_1(M)) = 3$ *and* $\pi_1(M) = *_m\mathbb{Z}$.

In short, the proofs of Propositions 4.2 and 4.3 are based on the fact that, for every crystallization Γ of a PL 4-manifold M, the associated triangulation $K(\Gamma)$ gives rise to a suitable handle-decomposition of M, which reflects the combinatorial properties of Γ.
First of all, we recall that every closed PL 4-manifold M admits a handle-decomposition

$$M = H^{(0)} \cup (H_1^{(1)} \cup \cdots \cup H_{r_1}^{(1)}) \cup (H_1^{(2)} \cup \cdots \cup H_{r_2}^{(2)}) \cup (H_1^{(3)} \cup \cdots \cup H_{r_3}^{(3)}) \cup H^{(4)}$$

where $H^{(0)} = \mathbb{D}^4$ and each p-handle $H_i^{(p)} = \mathbb{D}^p \times \mathbb{D}^{4-p}$ $(1 \leq p \leq 4,\ 1 \leq i \leq r_p)$ is endowed with an an embedding (called *attaching map*) $f_i^{(p)} : \partial\mathbb{D}^p \times \mathbb{D}^{4-p} \to \partial(H^{(0)} \cup \ldots (H_1^{(p-1)} \cup \cdots \cup H_{r_{p-1}}^{(p-1)}))$.

In particular, for any crystallization Γ of a PL 4-manifold M and for any partition $\{\{i,j,k\},\{r,s\}\}$ of Δ_4, then M admits a decomposition of type $M = N(i,j,k) \cup_\phi$

$N(r,s)$, where $N(i,j,k)$ (resp. $N(r,s)$) denotes a regular neighbourhood of the subcomplex $K(i,j,k)$ (resp. $K(r,s)$) of $K(\Gamma)$ generated by the vertices labelled $\{i,j,k\}$ (resp. $\{r,s\}$) and ϕ is a boundary identification.

The hypotheses assumed about regular genus in most of the cases of the above statements imply the associated handle-decomposition to lack in 2-handles, since the subcomplex $K(i,j,k)$ collapses to a graph; this fact allows to recognize the 4-manifold M as a connected sum of copies of $\mathbb{S}^1 \otimes \mathbb{S}^3$, by means of an important result by Montesinos [45] ensuring the uniqueness of the boundary identification of two handlebodies (see the quoted papers for details).

Under different assumptions, the associated handle-decomposition contains exactly one 2-handle and no 3-handle (i.e.: $K(i,j,k)$ consists of two triangles with common boundary, possibly with some more "free edges", in bijection with the "free edges" constituting $K(r,s)$); so, the attachment of the unique 2-handle has to give rise to a spherical boundary, and hence a $\mathbb{CP}^2$ component is surely obtained, via a well-known result by Gordon-Luecke about surgery on knots.

Remark 1 Note that the classification of orientable (resp. non-orientable) 4-manifolds with regular genus four (risp. three) is given in [29, Theorem 2] (resp. [29, Theorem 4]) only up to TOP-homeomorphism[8], while the classification of non-orientable 4-manifolds with regular genus four is not known.

However, Proposition 4.3 yields the following partial results, concerning the PL classification of a PL 4-manifold M with $rk(\pi_1(M)) = m$:

- if $\mathcal{G}(M) = 3$ and $m = 3$, then M is PL-homeomorphic to $\#_3(\mathbb{S}^1 \otimes \mathbb{S}^3)$;
- if $\mathcal{G}(M) = 4$, $m = 2$ and the fundamental group of M is free, then M is PL-homeomorphic to $\mathbb{CP}^2 \#_2(\mathbb{S}^1 \otimes \mathbb{S}^3)$.

Moreover, no PL 4-manifold exists with $\mathcal{G}(M) = 3$ (resp. $\mathcal{G}(M) = 4$) and $m \in \{0,2\}$ (resp. $m = 3$).

In Subsection 5.3 we will "almost complete"[9] the PL classification up to regular genus four, within the class of PL 4-manifolds admitting semi-simple crystallizations.

4.2. Catalogues of crystallizations

By Proposition 2.2, the generation of catalogues of crystallizations of PL 4-manifolds with a fixed number of vertices $2p$ requires the prior generation and recognition of all gems (with $2p$ vertices) representing the 3-sphere. In fact, any order $2p$ crystallization of a PL 4-manifold can be obtained from such a gem by adding the 4-coloured edges.

[8]Recall that, in virtue of Theorem 3.5, the classification of simply-connected 4-manifolds in category TOP is now trivial, up to regular genus 43.

[9]Only the case $\mathcal{G}(M) = 4$ and $rk(\pi_1(M)) = 2$, with not free fundamental group, remains open.

However, this kind of procedure is very intensive even for a low number of vertices and requires large computing resources. A way to face this problem is to find combinatorial configurations in the graphs, which can be eliminated without changing the manifold. Examples of such configurations are dipoles and ρ-pairs.

As pointed out in Section 2 the restriction of the catalogues to rigid dipole-free crystallizations does not affect their completeness.

Furthermore, an efficient catalogue must not contain crystallizations, whose associated triangulations coincide. This problem has been solved by associating to each crystallization Γ its *code*, i.e. a numerical string which identifies Γ up to *colour-isomorphisms* (i.e. isomorphisms of graphs which preserve colours up to a permutation of Δ_4: see [41], [24]).

Catalogues of (rigid dipole-free) crystallizations of PL 4-manifolds have been generated up to 20 vertices and the represented manifolds have been completely classified up to 18 vertices.

In order to describe the algorithms for generation and classification of these catalogues, we need first to fix some notations.

For each $p \geq 1$, let $\mathcal{C}^{(2p)}$ (resp. $\tilde{\mathcal{C}}^{(2p)}$) denote the catalogue of all not colour-isomorphic rigid dipole-free bipartite (resp. non-bipartite) crystallizations of 4-manifolds with $2p$ vertices.

Note that if $\Gamma \in \mathcal{C}^{(2p)} \cup \tilde{\mathcal{C}}^{(2p)}$, then $\Gamma_{\hat{4}}$ is a (not necessarily contracted, nor rigid) 4-coloured graph representing $\mathbb{S}^3$ and lacking in ρ-pairs involving three bicoloured cycles. Let $S^{(2p)}$ denotes the set of such 4-coloured graphs[10]: the generation of $\mathcal{C}^{(2p)}$ and $\tilde{\mathcal{C}}^{(2p)}$ is performed by adding 4-coloured edges to the elements of $S^{(2p)}$, in all possible ways so as to obtain crystallizations of 4-manifolds. Duplicates are then eliminated by comparing their codes.

Actually, not all attachments must be tried: the condition of representing a manifold imposes combinatorial restrictions on the set of possible "uncompleted" graphs (i.e. graphs obtained from an element of $S^{(2p)}$ by attaching less than p 4-coloured edges).

These conditions allow to implement a branch and bound technique to prune the tree of possible attachments (see [44] for details) and reduce considerably both the computation time and the size of the resulting catalogues[11].

Table 1 shows information about the catalogues $\mathcal{C}^{(2p)}$ and $\tilde{\mathcal{C}}^{(2p)}$ ($p \leq 10$), which have been obtained in [20] by the above algorithm.

[10] $S^{(2p)}$ is constructed by a suitable adaptation of the 3-dimensional generation algorithm; recognition of the 3-sphere is performed by comparison with the 3-dimensional catalogue.

[11] The generation algorithm has been implemented through a parallelization strategy ([44]); it ran on CINECA's high-performance clusters due to the opportunities granted by an Italian Supercomputing Resource Allocation (ISCRA) project.

2p	2	4	6	8	10	12	14	16	18	20
$\#S^{(2p)}$	1	0	2	9	39	400	5.255	95.870	1.994.962	45.654.630
$\#\ \mathcal{C}^{(2p)}$	1	0	0	1	0	0	1.109	4.511	44.803	47.623.129
$\#\ \tilde{\mathcal{C}}^{(2p)}$	0	0	0	0	0	0	0	1	0	0

Table 1

Note that the unique rigid dipole-free crystallization of $\mathcal{C}^{(2)}$ (resp. of $\mathcal{C}^{(8)}$) is the standard crystallization of $\mathbb{S}^4$ (resp. $\mathbb{CP}^2$: see [36]), while the unique non-bipartite rigid dipole-free crystallization appearing up to 20 vertices is the standard one of $\mathbb{RP}^4$ with 16 vertices ([37]).

In order to face the problem of classification of the PL-manifolds appearing in catalogues of crystallizations, in [20] it is described a heuristic procedure based on combinatorial moves on graphs which do not change ("up to handles") the represented manifold and preserve the properties of rigidity and absence of dipoles.

We point out that all concepts and results involved in the procedure hold in each dimension $n \geq 3$; therefore the whole classifying algorithm is introduced in [20] for general PL n-manifolds. Nevertheless, for sake of simplicity, in the present survey paper we will describe it only in the 4-dimensional setting.

Let us call *admissible* any sequence of combinatorial moves which transforms a rigid dipole-free crystallization of a PL 4-manifold M, into a rigid dipole-free crystallization of a PL 4-manifold M' such that $M \cong_{PL} M' \#_h (\mathbb{S}^1 \otimes \mathbb{S}^3)$ $(h \geq 0)$.

Moreover, for any rigid dipole-free crystallization Γ of M and any admissible sequence ϵ, let $\theta_\epsilon(\Gamma)$ denote the (rigid dipole-free) crystallization of M' obtained by applying the admissible sequence ϵ to Γ.

Given a list X of rigid dipole-free crystallizations and a set $\mathcal{S}$ of admissible sequences, it is then possible to subdivide X into equivalence classes with regard to $\mathcal{S}$ by defining the class of $\Gamma \in X$ with respect to $\mathcal{S}$ as:

$$cl_{\mathcal{S}}(\Gamma) = \{\Gamma' \in X \ / \ \exists \epsilon, \epsilon' \in \mathcal{S},\ \theta_\epsilon(\Gamma) \text{ and } \theta_{\epsilon'}(\Gamma') \text{ have the same code}\}$$

Therefore, given $\Gamma, \Gamma' \in X$, if $cl_{\mathcal{S}}(\Gamma) = cl_{\mathcal{S}}(\Gamma')$, then there exist $h, k \in \mathbb{N} \cup \{0\}$ such that $|K(\Gamma)| \cong_{PL} M \#_h (\mathbb{S}^1 \otimes \mathbb{S}^3)$ and $|K(\Gamma')| \cong_{PL} M \#_k (\mathbb{S}^1 \otimes \mathbb{S}^3)$.

Obviously, no choice of $\mathcal{S}$ can ensure "a priori" that the above equivalence classes coincide, even up to handles, with the PL-equivalence classes of the represented 4-manifolds.

However, [20] shows the existence of a suitable set of admissible moves on 5-coloured graphs which are sufficient to classify all PL 4-manifolds admitting a crystallization with at most 18 vertices, in full analogy with what already proved in dimension three, where a similar set has been detected, yielding the classification of all 3-manifolds admitting a crystallization with at most 30 vertices.

It is well-known that insertion or elimination of a dipole (*dipole move*) preserves the PL structure of the represented manifold in any dimension (see [30] for details).

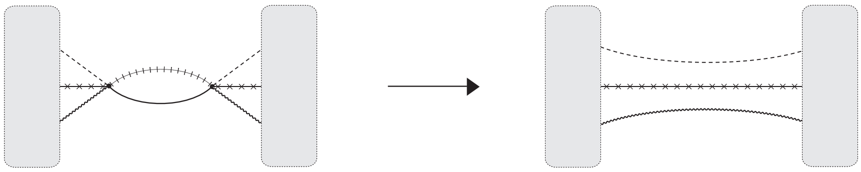


Figure 1: dipole move

Furthermore, a combinatorial move called *ρ-pair switching*, which is shown in Figure 4.2, allows to eliminate ρ-pairs.

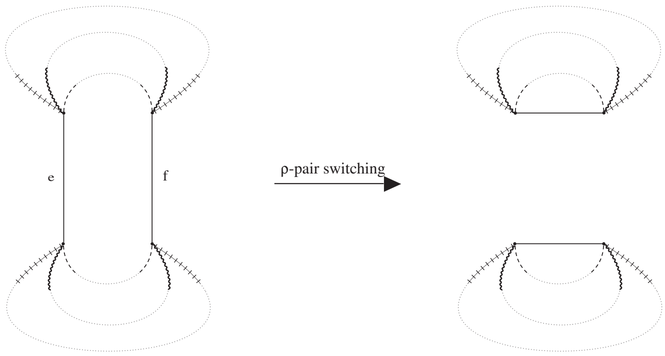


Figure 2: ρ-pair switching

The effect of ρ-pair switching on crystallizations is explained by the following result:

Proposition 4.4 [5] *Let Γ be a crystallization of a PL 4-manifold M and let Γ' be obtained by switching a ρ-pair (e, f) in Γ. Suppose that e and f share $h \geq 3$ bicoloured cycles of Γ. Then:*

(a) if $h = 3$, Γ' is a gem of M, too;

(b) if $h = 4$, Γ' is a gem of a PL 4-manifold M' such that $M \cong_{PL} M' \#(\mathbb{S}^1 \otimes \mathbb{S}^3)$.

Unfortunately, some moves which turned out to be powerful for the classification of 3-manifolds are not available in dimension greater than three. Therefore, the algorithm described in [20] makes use of other moves which were introduced by Lins and Mulazzani in [42].

Let Γ be a gem of a PL 4-manifold M.

- A *blob move* is the insertion or cancellation of a 4-dipole.
- A *t-flip* is the switching of a pair (e, f) of equally coloured edges which are both incident to an h-dipole $(1 \le h \le 3)$. An *s-flip* is the inverse move, i.e. the switching of a pair (e, f) of equally coloured edges where either e or f belongs to an h-dipole, which becomes an $(h-1)$-dipole after the transformation. A *flip* is either an s- or a t-flip.

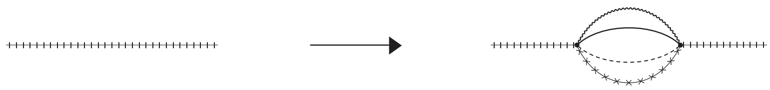


Figure 3: blob move

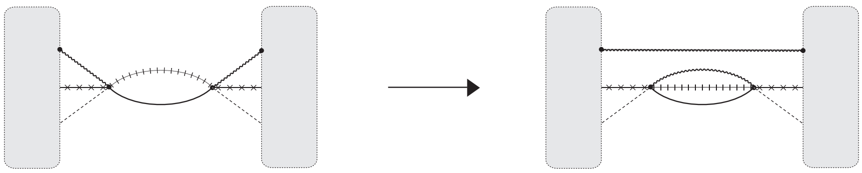


Figure 4: flip move

Flip and blob moves on a gem do not change the represented manifold as proved in [42, Proposition 3].

On the other hand, we point out that, even if two crystallizations are known to represent the same manifold, there is no algorithmic procedure to determine a sequence of blob and flip moves connecting them, nor an upper bound to the number of moves to be performed.

In order to define the set of admissible moves $\bar{\mathcal{S}}$ which have been chosen for the heuristic procedure, let us introduce some definitions and notations.

Given an order $2p$ 5-coloured graph Γ there is a natural ordering of its vertices induced by the *rooted numbering algorithm* generating its code (see [24]); so $V(\Gamma) = \{v_1, \dots, v_{2p}\}$ may be assumed.

If Γ is a rigid dipole-free crystallization of a PL 4-manifold, given $i \in \mathbb{N}_{2p} = \{1, \dots, 2p\}$, $c \in \Delta_4$, a 4-tuple $\mathbf{x} = (x_1, \dots, x_4)$ with $x_i \in \mathbb{N}_{2p}$ and a permutation τ of $\hat{c} = \Delta_4 - \{c\}$, we denote by $\theta_{i,c,\mathbf{x},\tau}(\Gamma)$ the rigid dipole-free crystallization obtained from Γ in the following way:

- insert a 4-dipole (= blob) over the c-coloured edge incident with v_i;
- for each $k \in \hat{c}$, consider, if exists, the s-flip on the pair of $\tau(k)$-coloured edges (e, f), where e belongs to the blob and f is incident to v_{x_k}; then perform the sequence of all possible s-flips of this type for increasing values of k;

- cancel dipoles and switch ρ-pairs in the resulting graph.

$\theta_{i,c,\mathbf{x},\tau}$ obviously defines an admissible sequence.

Finally, $\bar{\mathcal{S}}$ is defined as the set of all sequences $\theta_{i,c,\mathbf{x},\tau}$, where $i \in \mathbb{N}_{2p}$, $c \in \Delta_4$, $\mathbf{x}$ is a 4-tuple of elements of $\mathbb{N}_{2p}$ and τ is a permutation of $\hat{c}$.

Remark 2 As already mentioned, the above moves, as well as the classification algorithm itself, are independent from dimension. As a consequence, the set $\bar{\mathcal{S}}$ can be defined and the partition into equivalence classes with respect to $\bar{\mathcal{S}}$ can be performed on any list of crystallizations of n-manifolds, in order to prove their PL-equivalence. See [20] (or Subsection 6.1) for examples of application of the above algorithm.

4.3. Classification results in PL category

In order to obtain PL classification results, the classification algorithm, with respect to the above defined set $\bar{\mathcal{S}}$, has been implemented in a C++ program - called "Γ4-*class*" and it has been applied to the catalogues $\mathcal{C}^{(2p)}$ and $\tilde{\mathcal{C}}^{(2p)}$ with $p \leq 9$ and to the subset of $\mathcal{C}^{(20)}$ consisting of crystallizations representing manifolds with $\beta_2 \leq 2$.

The application of Γ4-*class* to the catalogue $\bigcup_{1\leq p\leq 9} \mathcal{C}^{(2p)}$ yielded the complete PL classification of the involved crystallizations as shown in the following proposition.

Proposition 4.5 [20] *There is a bijective correspondence between the set of equivalence classes of* $\bigcup_{1\leq p\leq 9} \left(\mathcal{C}^{(2p)} \cup \tilde{\mathcal{C}}^{(2p)}\right)$ *with respect to* $\bar{\mathcal{S}}$ *and the set of the represented PL 4-manifolds. Moreover, all PL 4-manifolds in the above catalogues are topologically distinct.*

By the above result and [20, Proposition 15], it has been obtained the PL classification of all orientable (resp. non-orientable) 4-manifolds with gem-complexity at most 8 (resp. at most 9), which is summarized in the following theorem. Note that the last statement appears in the present survey in a stronger form than the original result in [20]: in fact, only recently program Γ4-*class* succeeded to prove that no PL 4-manifold M with $k(M) = 9$ has $\beta_2(M) \leq 2$ (i.e. all 20 vertices crystallizations with $\beta_2(M) \leq 2$ belongs to the same class of a crystallization with few vertices).

Theorem 4.6 *Let M be a PL 4-manifold. Then:*

- $k(M) = 0 \iff$ *M is PL-homeomorphic to* $\mathbb{S}^4$;
- $k(M) = 3 \iff$ *M is PL-homeomorphic to* $\mathbb{CP}^2$;
- $k(M) = 4 \iff$ *M is PL-homeomorphic to either* $\mathbb{S}^1 \times \mathbb{S}^3$ *or* $\mathbb{S}^1 \tilde{\times} \mathbb{S}^3$;
- $k(M) = 6 \iff$ *M is PL-homeomorphic to either* $\mathbb{S}^2 \times \mathbb{S}^2$ *or* $\mathbb{CP}^2 \# \mathbb{CP}^2$ *or* $\mathbb{CP}^2 \# (-\mathbb{CP}^2)$;

- $k(M) = 7 \iff$ *M is PL-homeomorphic to either* $\mathbb{RP}^4$ *or* $\mathbb{CP}^2\#(\mathbb{S}^1 \times \mathbb{S}^3)$ *or* $\mathbb{CP}^2\#(\mathbb{S}^1 \widetilde{\times} \mathbb{S}^3)$;
- $k(M) = 8 \iff$ *M is PL-homeomorphic to either* $\#_2(\mathbb{S}^1 \times \mathbb{S}^3)$ *or* $\#_2(\mathbb{S}^1 \widetilde{\times} \mathbb{S}^3)$.

Moreover:

- *no PL* 4*-manifold M exists with* $k(M) \in \{1, 2, 5\}$;
- *no exotic PL* 4*-manifold exists, with* $k(M) \leq 8$;
- *any PL* 4*-manifold M with* $k(M) = 9$ *is simply connected with second Betti number* $\beta_2(M) = 3$.

As a consequence of the (partial) analysis of the 4-dimensional crystallization catalogue $\bigcup_{1 \leq p \leq 10} \mathcal{C}^{(2p)}$, together with a suitable application of the classification program Γ4-*class*, in [20] it is also proved the existence of a rigid crystallization of $\mathbb{S}^4$, with 20 vertices and, apart from the standard order-two crystallization, it is the only rigid dipole-free crystallization of $\mathbb{S}^4$ up to 20 vertices.

5. 4-manifolds admitting simple and semi-simple crystallizations

5.1. Simple and semi-simple crystallizations

The notion of *simple crystallization* of a (simply-connected) PL 4-manifold was introduced in [7] and investigated in [21]; further, in [6], it was extended to the not simply-connected case, by introducing the concept of *semi-simple crystallization* of a PL 4-manifold.

Definition 3 A crystallization Γ of a PL 4-manifold M is called a *semi-simple crystallization of type m* if the 1-skeleton of the associated coloured triangulation contains exactly $m+1$ 1-simplices for each pair of 0-simplices, where m is the rank of the fundamental group of M.

Semi-simple crystallizations of type 0 are called *simple crystallizations*: the 1-skeleton of their associated coloured triangulation equals the 1-skeleton of a single 4-simplex.

Remark 3 In virtue of the bijection between 1-simplices of $K(\Gamma)$ and residues of Γ with three colours (see Section 2), the above definition may be re-stated in combinatorial terms, for any crystallization Γ of a PL 4-manifold M with $rk(\pi_1(M)) = m$:

$$\Gamma \text{ is a semi-simple crystallization} \iff g_{ijk} = m + 1, \ \forall i, j, k \in \Delta_4.$$

In particular:

$$\Gamma \text{ is a simple crystallization} \iff g_{ijk} = 1, \ \forall i, j, k \in \Delta_4.$$

As a direct consequence of the proof of Theorem 3.1 (inequality concerning gem-complexity), a characterization of PL 4-manifolds admitting simple/semi-simple crystallizations easily follows, involving the relation between the invariant gem-complexity and the Euler characteristic.

Theorem 5.1 [21, 6] *A PL 4-manifold M admits semi-simple crystallizations of type m if and only if $k(M) = 3\chi(M)+10m-6$, where m is the rank of the fundamental group of M.*
In particular: A simply-connected PL 4-manifold M admits simple crystallizations if and only if $k(M) = 3\chi(M) - 6$.

In [21] (resp. in [6]) simple (resp. semi-simple) crystallizations are proved to be "minimal" both with respect to the invariant gem-complexity and with respect to the invariant regular genus. Here, we present the generalized result concerning semi-simple crystallizations, which contextually yields a lot of details about their combinatorial structure.

Theorem 5.2 [6] *Let M be a PL 4-manifold with $rk(\pi_1(M)) = m$. If M admits semi-simple crystallizations, then:*

$$k(M) = 3\chi(M) + 10m - 6;$$
$$\mathcal{G}(M) = 2\chi(M) + 5m - 4;$$
$$k(M) = \frac{3\mathcal{G}(M) + 5m}{2}.$$

Moreover, for any semi-simple crystallization Γ of M,

- $\rho_\epsilon(\Gamma) = \mathcal{G}(M) = 2\chi(M) + 5m - 4$ *for any cyclic permutation ϵ of Δ_4;*
- $\#V(\Gamma) = 2(k(M) + 1) = 6\chi(M) + 20m - 10;$
- $g_{i,j} = \chi(M) + 4m - 1$ *for any pair $i, j \in \Delta_4$;*
- $\rho_\epsilon(\Gamma_{\hat{i}}) = \frac{\mathcal{G}(M)-m}{2} = \chi(M) + 2m - 2$ *for any cyclic permutation ϵ of Δ_4 and for any color $i \in \Delta_4$.*

In the simply-connected case, the characterization of PL 4-manifolds admitting simple crystallizations via gem-complexity may be performed with respect to the second Betti number, as proved for the first time in [21, Theorem 1.1].

Theorem 5.3 [21] *Let M be a simply-connected PL 4-manifold. Then:*

$$M \text{ admits a simple crystallization iff } k(M) = 3\beta_2(M).$$

Moreover, if M admits simple crystallizations, then $\mathcal{G}(M) = 2\beta_2(M)$.

Remark 4 It is not difficult to prove that, if $\pi_1(M)$ is assumed to be trivial, then equality $\mathcal{G}(M) = 2\beta_2(M)$ implies the existence of a crystallization Γ of M and a permutation ε of Δ_4 so that $\rho_\varepsilon(\Gamma) = 2\beta_2(M)$ and $g_{\varepsilon_i\varepsilon_{i+2}\varepsilon_{i+3}} = 1\ \forall i \in \Delta_4$. However, in general, this does not imply that Γ is simple, since at least one $g_{\varepsilon_i\varepsilon_{i+1}\varepsilon_{i+2}} > 1$ may occur. For example, the analysis of the catalogue of rigid dipole-free order 16 crystallizations shows that all of them satisfy relation $\mathcal{G}(M) = 2\beta_2(M)$, while $g_{rst} = 2$ for exactly one triple $\{r,s,t\} \subset \Delta_4$ and $g_{ijk} = 1\ \forall \{i,j,k\} \neq \{r,s,t\}$.

Let us conclude the subsection by pointing out that, in the particular case of a simple crystallization, any associated handle-decomposition (see Subsection 4.1) turns out to be a so-called *special handle decomposition*, i.e. a handle-decomposition lacking in 1-handles and 3-handles (see [43, Section 3.3]). Hence, PL 4-manifolds admitting simple crystallizations may be identified by a (not dotted) framed link.[12]

Proposition 5.4 [21] *Let M be a (simply-connected) PL 4-manifold admitting simple crystallizations. Then, M admits a special handle-decomposition.*

In fact, with the same notations used in Subsection 4.1, we can notice that, if Γ is a simple crystallization, then for any partition $\{\{i,j,k\},\{r,s\}\}$ of Δ_4 the decomposition $M = N(i,j,k) \cup_\phi N(r,s)$ is of "standard" type: $K(r,s)$ consists of exactly one 1-simplex, while $K(i,j,k)$ consists of g_{rs} 2-simplices, all having the same boundary. Hence, $N(r,s) \cong_{PL} \mathbb{D}^4 = H^{(4)}$ trivially follows, while $N(i,j,k) = H^{(0)} \cup (H^{(2)}_1 \cup \cdots \cup H^{(2)}_{g_{rs}-1})$ holds, where $H^{(0)} = \mathbb{D}^4$ is a "small" regular neighbourhood of one (arbitrarily fixed) 2-simplex of $K(i,j,k)$ and the 2-handles are represented by the regular neighbourhoods of the remaining 2-simplices of $K(i,j,k)$.

Remark 5 Note that the existence of a special handlebody decomposition is related to Kirby problem n. 50: *"Does every simply-connected closed 4-manifold have a handlebody decomposition without 1-handles? Without 1- and 3-handles?"*.

5.2. Computing regular genus and gem-complexity for a huge class of PL 4-manifolds

The definition of graph connected sum (see Section 2) implies that the class of PL 4-manifolds admitting simple/semi-simple crystallization is closed under connected sum.

Proposition 5.5 [7] *Let M and M' be two PL 4-manifolds admitting semi-simple crystallizations. Then, $M\#M'$ admits semi-simple crystallizations, too.*
In particular, if both M and M' admit simple crystallizations, then $M\#M'$ admits simple crystallizations, too.

[12] See [15] for relationships between crystallization theory and dotted framed link representation for PL 4-manifolds.

It is easy to check that the well-known minimal (order 10) crystallizations of $\mathbb{S}^1 \times \mathbb{S}^3$ and $\mathbb{S}^1 \widetilde{\times} \mathbb{S}^3$, as well as the minimal (order 16) crystallization of $\mathbb{RP}^4$, are semi-simple crystallizations of type 1, while the minimal (order 2) crystallization of $\mathbb{S}^4$, the minimal (order 8) crystallization of $\mathbb{CP}^2$, the minimal (order 14) crystallization of $\mathbb{S}^2 \times \mathbb{S}^2$ are simple crystallizations. Moreover, in [7] a simple crystallization of the $K3$-surface is produced.

As a consequence, in virtue of the additivity of semi-simple crystallizations (Proposition 5.5), we have that all simply-connected PL 4-manifolds of "standard type" (see [7]) admit semi-simple crystallizations, as well as all PL 4-manifolds involved in the existing crystallization catalogues (see [20], or Section 4).

Proposition 5.6 [7] *Each PL 4-manifold with gem-complexity less than nine admits semi-simple crystallizations.*

By direct analysis of the existing 4-dimensional crystallization catalogues (see [20] or Section 4), we can compute how many simple/semi-simple crystallizations exist, for some significant PL 4-manifolds.

Proposition 5.7

- $\mathbb{S}^4$ *and* $\mathbb{CP}^2$ *admit a unique simple crystallization;*
- $\mathbb{S}^1 \times \mathbb{S}^3$, $\mathbb{S}^1 \widetilde{\times} \mathbb{S}^3$ *and* $\mathbb{RP}^4$ *admit a unique semi-simple crystallization (of type 1);*
- $\mathbb{S}^2 \times \mathbb{S}^2$ *admits exactly* 267 *simple crystallizations;*
- $\mathbb{CP}^2 \# \mathbb{CP}^2$ *admits exactly* 583 *simple crystallizations;*
- $\mathbb{CP}^2 \# (-\mathbb{CP}^2)$ *admits exactly* 258 *simple crystallizations.*

From Theorem 5.2 and Proposition 5.5 it is easy to deduce the additivity of both the invariants regular genus and gem-complexity under connected sum, within the class of PL 4-manifolds admitting semi-simple crystallizations (in particular: simple crystallizations).

Theorem 5.8 [21, 7] *Let M and M' be two PL 4-manifolds admitting semi-simple crystallizations. Then:*

$$k(M \# M') = k(M) + k(M') \quad \textit{and} \quad \mathcal{G}(M \# M') = \mathcal{G}(M) + \mathcal{G}(M').$$

As a consequence, we obtain the computation of both invariants for a huge class of PL 4-manifolds.

Proposition 5.9 *Let*

$$M \cong_{PL} (\#_p \mathbb{CP}^2)\#(\#_{p'}(-\mathbb{CP}^2))\#(\#_q(\mathbb{S}^2 \times \mathbb{S}^2))\#(\#_r(\mathbb{S}^1 \otimes \mathbb{S}^3))\#(\#_s \mathbb{RP}^4)\#(\#_t K3),$$

with $p, p', q, r, s, t \geq 0$. *Then,*

$$k(M) = 3(p+p'+2q+22t)+4r+7s \quad \text{and} \quad \mathcal{G}(M) = 2(p+p'+2q+22t)+r+3s.$$

5.3. Classification via regular genus of PL 4-manifolds admitting simple/semi-simple crystallizations

In the particular case of PL 4-manifolds admitting semi-simple crystallizations, some classification results via regular genus may be added to the general ones summarized in Subsection 4.1. The first one concerns the case of the first Betti number equal to one.

Proposition 5.10 [7] *Let M be an orientable PL 4-manifold with $\pi_1(M) = *_m\mathbb{Z}$ and $\beta_2(M) = 1$. If M admits semi-simple crystallizations, then M is PL-homeomorphic to $\#_m(\mathbb{S}^1 \times \mathbb{S}^3)\#\mathbb{CP}^2$.*

Moreover, the relation between regular genus and gem-complexity for PL 4-manifolds admitting semi-simple crystallizations yields new results regarding the PL classification up to regular genus four, within the class of PL 4-manifolds admitting a semi-simple crystallization. As already pointed out in Subsection 4.1, only the case $\mathcal{G}(M) = 4$ and $rk(\pi_1(M)) = 2$, with not free fundamental group, remains open.

Proposition 5.11 *Let M be a PL 4-manifold with $rk(\pi_1(M)) = m$ which admits semi-simple crystallizations. Then:*

(a) if $\mathcal{G}(M) = 3$ and $m = 1$, then M is PL-homeomorphic to one of the following PL 4-manifolds: $\mathbb{CP}^2\#(\mathbb{S}^1 \times \mathbb{S}^3)$, $\mathbb{CP}^2\#(\mathbb{S}^1 \widetilde{\times} \mathbb{S}^3)$, $\mathbb{RP}^4$;

(b) if $\mathcal{G}(M) = 3$ and $m = 3$, then M is PL-homeomorphic to $\#_3(\mathbb{S}^1 \otimes \mathbb{S}^3)$;

(c) if $\mathcal{G}(M) = 4$ and $m = 0$, then M is PL-homeomorphic to one of the following PL 4-manifolds: $\mathbb{CP}^2\#\mathbb{CP}^2$, $\mathbb{S}^2 \times \mathbb{S}^2$, $\mathbb{CP}^2\#(-\mathbb{CP}^2)$;

(d) if $\mathcal{G}(M) = 4$, $m = 2$ and the fundamental group of M is free, then M is PL-homeomorphic to $\mathbb{CP}^2\#_2(\mathbb{S}^1 \otimes \mathbb{S}^3)$.

Moreover, if $\mathcal{G}(M) = 3$ (resp. $\mathcal{G}(M) = 4$), the cases $m \in \{0, 2\}$ (resp. $m \in \{1, 3\}$) cannot appear for PL 4-manifolds admitting semi-simple crystallizations.

Proof. First of all, note that, by Remark 1, only statements (a) and (c) require the hypothesis about semi-simple crystallizations, and have to be proved.

By Theorem 5.2, if M is a PL 4-manifold admitting semi-simple crystallizations, relation $k(M) = \frac{3\mathcal{G}(M)+5m}{2}$ holds. Then: in case (a) $k(M) = 7$ follows, while in case (c) $k(M) = 6$ follows. Hence, statements (a) and (c) are consequence of the PL classification of all the (orientable and non-orientable) PL 4-manifolds up to gem-complexity 8, obtained in [20] via analysis of the related crystallization catalogues (see Theorem 4.6).

Moreover, the relation $k(M) = \frac{3\mathcal{G}(M)+5m}{2}$ easily implies that the cases $\mathcal{G}(M) = 3$ (resp. $\mathcal{G}(M) = 4$) and $m \in \{0, 2\}$, (resp. $m \in \{1, 3\}$) are impossible, if M admits semi-simple crystallizations. $\square$

General conditions excluding the existence of semi-simple crystallizations are also obtained from their combinatorial properties (Theorem 5.2).

Proposition 5.12 [7] *No PL* 4*-manifold* M *with* $\mathcal{G}(M) - rk(\pi_1(M))$ *odd admits semi-simple crystallizations.*
In particular: no simply-connected PL 4*-manifold* M *with odd regular genus admits simple crystallizations.*

6. Toward further developments

6.1. Different PL structures on the same TOP 4-manifold, and related problems

As it is well-known, up to now there is no classification of smooth structures on any given smoothable topological 4-manifold; on the other hand, finding non-diffeomorphic smooth structures on the same closed simply-connected topological manifold has long been an interesting problem.

Our hope is that further advances in the generation and classification of crystallization catalogues for PL 4-manifolds according to gem-complexity (see the algorithm described in [20, Section 3] and briefly summarized in Subsection 4.2) could produce examples of non-equivalent PL structures on the same TOP 4-manifold.

For example, the characterization of PL 4-manifolds admitting semi-simple crystallizations by means of gem-complexity (Theorem 5.1) has the following consequence about possible different PL structures on the same TOP 4-manifold.

Proposition 6.1 *Let* M *and* M' *be two PL* 4*-manifolds, with* $M \cong_{TOP} M'$ *and* $M \ncong_{PL} M'$*. If both* M *and* M' *admit semi-simple crystallization, then* $k(M) = k(M')$*.*

Remark 6 In particular, in the simply-connected case, the catalogue of all rigid dipole-free crystallizations of PL 4-manifolds up to a fixed gem-complexity k must contain all simple crystallizations of PL 4-manifolds whose second Betti number does not exceed $\frac{k}{3}$. Hence, the existing catalogue up to gem-complexity 9 ([20]) presents all simple crystallizations of PL 4-manifolds M with $\beta_2(M) \leq 3$.

On the other hand, in [1], infinitely many PL 4-manifolds TOP-homeomorphic but not PL-homeomorphic to $\mathbb{CP}^2\#_2(-\mathbb{CP}^2)$ are proved to exist. Hence, if at least one among them admits simple crystallizations, they have to appear in the catalogue of order 20 crystallizations, whose analysis is currently underway.

More generally, the existence of simple/semi-simple crystallizations may be related with known results and open problems about exotic structures on "standard" simply-connected PL 4-manifolds and on non-orientable PL 4-manifolds (see for example [1], [2] and [10]), by taking into account also the (obvious) finiteness property of gem-complexity. Items (a), (d), (e) of the following proposition are due to [20, 21], while items (b) and (c) are new.

Proposition 6.2

(a) Let M be $\mathbb{S}^4$ or $\mathbb{CP}^2$ or $\mathbb{S}^2 \times \mathbb{S}^2$ or $\mathbb{CP}^2\#\mathbb{CP}^2$ or $\mathbb{CP}^2\#(-\mathbb{CP}^2)$; if an exotic PL structure on M exists, then the corresponding PL-manifold does not admit simple crystallizations.

(b) Let $\bar{M}$ be a PL 4-manifold TOP-homeomorphic but not PL-homeomorphic to $\mathbb{RP}^4$; then, $\bar{M}$ does not admit semi-simple crystallizations.

(c) Let M be $\mathbb{S}^1 \otimes \mathbb{S}^3$ or $\mathbb{CP}^2\#(\mathbb{S}^1 \otimes \mathbb{S}^3)$ or $\#_2(\mathbb{S}^1 \otimes \mathbb{S}^3)$; if an exotic PL structure on M exists, then the corresponding PL-manifold does not admit semi-simple crystallizations.

(d) Let $\bar{M}$ be a PL 4-manifold TOP-homeomorphic but not PL-homeomorphic to $\mathbb{CP}^2\#_2(-\mathbb{CP}^2)$; then, either $\bar{M}$ does not admit simple crystallizations, or $\bar{M}$ admits an order 20 simple crystallization.

(e) Let $r \in \{3,5,7,9,11,13\} \cup \{r = 4n-1 \ / \ n \geq 4\} \cup \{r = 4n-2 \ / \ n \geq 23\}$; then, infinitely many simply-connected PL 4-manifolds with $\beta_2 = r$ do not admit simple crystallizations.

As pointed out by [7, Corollary 8.3], the existence of pairs of simply-connected "standard" PL 4-manifolds which are TOP-homeomorphic but not PL-homeomorphic has a consequence involving simple crystallizations, too.

Proposition 6.3 [7] *A pair of simple crystallizations Γ, Γ' exists, such that*

$$|K(\Gamma)| \cong_{TOP} |K(\Gamma')| \quad \textit{but} \quad |K(\Gamma)| \ncong_{PL} |K(\Gamma')|.$$

In fact, by Kronheimer and Mrowka (see [40]), it is known that $K3\#(-\mathbb{CP}^2) \cong_{TOP} \#_3(\mathbb{CP}^2)\#_{20}(-\mathbb{CP}^2)$ but $K3\#(-\mathbb{CP}^2) \not\cong_{PL} \#_3(\mathbb{CP}^2)\#_{20}(-\mathbb{CP}^2)$. Hence, to obtain a pair of simple crystallizations proving the above statement, it is sufficient to make graph connected sums of suitable copies of the (known) simple crystallizations of $K3$ and $\mathbb{CP}^2$ (see Subsection 5.2).

Remark 7 If Γ' and Γ'' is a pair of simple crystallizations of M' and M'' respectively, with $M' \cong_{TOP} M''$ and $M' \not\cong_{PL} M''$ (for example, the pair obtained via Kronheimer and Mrowka's result), then the associated contracted pseudo-triangulations $K' = K(\Gamma')$ and $K'' = K(\Gamma'')$ are such that $N'(1,3) = N''(1,3) = \mathbb{D}^4$, while $N'(0,2,4)$ and $N''(0,2,4)$ both consist of $\beta_2(M') = \beta_2(M'')$ triangles with the same boundary (see Proposition 5.4). Hence, the same handle-decomposition is induced, and also with the same intersection form. However, the fact that $M' \not\cong_{PL} M''$ proves that, if $\beta_2(M') = \beta_2(M'') \geq 2$, it is not possible to identify the framed link associated to the 2-handles, despite what happens when $\beta_2(M') = \beta_2(M'') = 1$ by means of Gordon-Luecke's result.

Now, the semi-simple case of Proposition 6.1, together with results by [6], enables to extend Proposition 6.3 to the non-simply-connected case.

Proposition 6.4 *A pair of semi-simple crystallizations (of type $m \geq 1$) Γ, Γ' exists, such that*

$$|K(\Gamma)| \cong_{TOP} |K(\Gamma')| \quad \textit{but} \quad |K(\Gamma)| \not\cong_{PL} |K(\Gamma')|.$$

Proof. By Kreck (see [39]), it is known that $\mathbb{RP}^4\#K3 \cong_{TOP} \mathbb{RP}^4\#_{11}(\mathbb{S}^2 \times \mathbb{S}^2)$ but $\mathbb{RP}^4\#K3 \not\cong_{PL} \mathbb{RP}^4\#_{11}(\mathbb{S}^2 \times \mathbb{S}^2)$. Hence, to obtain a pair of semi-simple crystallizations (of type 1) proving the above statement, it is sufficient to make graph connected sums of the (unique) semi-simple crystallization of $\mathbb{RP}^4$ and either a simple crystallization of $K3$ or eleven copies of a simple crystallization of $\mathbb{S}^2 \times \mathbb{S}^2$ (see Subsection 5.2). □

Moreover, we point out that the program $\Gamma 4$-*class*, performing automatic recognition of PL-homeomorphic 4-manifolds, could be an useful tool to approach open problems related to different triangulations of the same TOP 4-manifold, which are conjectured to represent the same PL 4-manifold.

For example, it is in progress its application to the open problem concerning the possible PL-equivalence of the 16- and 17-vertices triangulations of the $K3$-surface obtained in [26] and [46] respectively.

Note that similar attempts to settle the conjecture are described in [7], [8] and [9]. However, the elementary moves involved in those procedures (namely, *edge-contraction* and/or *bistellar moves*) are different from those used by our program (i.e. flips, blobs, ρ-pair switchings and dipole eliminations). Hence, it is possible that one sequence succeeds when the others fail, or viceversa, with equal computational time employed.

6.2. Additivity of regular genus and related problems

It is easy to check that the relation $\mathcal{G}(M\#M') \leq \mathcal{G}(M) + \mathcal{G}(M')$ can be stated for all PL n-manifolds by direct estimation of $\mathcal{G}(M\#M')$ on the gem $\Gamma\#\Gamma'$, when Γ, Γ' are assumed to be gems of M, M' realizing the regular genus of the represented n-manifolds. Moreover, the additivity of regular genus under connected sum has been conjectured[13], and the associated (open) problem is significant, at least in the orientable case, and especially in dimension four.

Conjecture 1 [31] *Let M_1^n, M_2^n be two closed (orientable) PL n-manifolds. Then,*

$$\mathcal{G}(M_1^n \# M_2^n) = \mathcal{G}(M_1^n) + \mathcal{G}(M_2^n).$$

In fact, it is easy to prove that the 4-dimensional case of Conjecture 1 implies the 4-dimensional Smooth Poincaré Conjecture, via the well-known Wall Theorem on homotopic 4-manifolds:

if Σ^4 is a homotopy sphere, then $\Sigma^4\#(\#_h(\mathbb{S}^2\times\mathbb{S}^2)) \cong \mathbb{S}^4\#(\#_h(\mathbb{S}^2\times\mathbb{S}^2)) \cong \#_h(\mathbb{S}^2\times\mathbb{S}^2)$, for a suitable non-negative integer h, and hence $\mathcal{G}(\Sigma^4\#(\#_h(\mathbb{S}^2\times\mathbb{S}^2))) = \mathcal{G}(\#_h(\mathbb{S}^2\times\mathbb{S}^2))$. Thus, the additivity of the regular genus would yield $\mathcal{G}(\Sigma^4) = 0$, i.e. $\Sigma^4 \cong \mathbb{S}^4$ (by Proposition 5.1, in case $n = 4$).

The following statement improves via Theorem 3.1 a double inequality concerning regular genus obtained in [38, Corollary 6.5].

Proposition 6.5 *For each closed PL 4-manifold M, with $rk(\pi_1(M)) = m$:*

$$2 - 2\mathcal{G}(M) \leq \chi(M) \leq 2 + \frac{\mathcal{G}(M)}{2} - \frac{5m}{2}.$$

In [38, Corollary 6.8], by means of the double inequality of [38, Corollary 6.5], two classes of (not necessarily orientable) PL 4-manifolds have been detected, for which additivity of regular genus holds. Now, by means of the improvement of Proposition 6.5, we can strictly enlarge the set of PL 4-manifolds for which additivity of regular genus is known to hold.

Proposition 6.6 *Let M_1, M_2 be two PL 4-manifolds, with $rk(\pi_1(M_i)) = m_i$ for each $i \in \{1, 2\}$.*

(a) If $\mathcal{G}(M_i) = 1 - \frac{\chi(M_i)}{2}$ for each $i \in \{1, 2\}$, then:

$$\mathcal{G}(M_1\#M_2) = \mathcal{G}(M_1) + \mathcal{G}(M_2) \quad \textit{and} \quad \mathcal{G}(M_1\#M_2) = 1 - \frac{\chi(M_1\#M_2)}{2}.$$

[13] Obviously, regular genus satisfies the additive property with respect to connected sum of closed 3-manifolds, via a classical result on Heegaard genus.

(b) If $\mathcal{G}(M_i) = 2\chi(M_i) + 5m_i - 4$ *for each* $i \in \{1, 2\}$, *then:*

$$\mathcal{G}(M_1\#M_2) = \mathcal{G}(M_1)+\mathcal{G}(M_2) \quad \text{and} \quad \mathcal{G}(M_1\#M_2) = 2\chi(M_1\#M_2)+5(m_1+m_2)-4.$$

As pointed out in [21], the class of PL 4-manifolds involved in Proposition 6.6(a) consists of connected sums of $\mathbb{S}^3$-bundles over $\mathbb{S}^1$: in fact, by the combinatorial properties of crystallizations in dimension 4, relation $\mathcal{G}(M) = 1 - \frac{\chi(M)}{2}$ implies $\rho_\epsilon(\Gamma_{\hat{i}}) = 0$ for each $i \in \Delta_4$, and $M \cong_{PL} \#_m(\mathbb{S}^1 \otimes \mathbb{S}^3)$ directly follows from the existence of (at least) an $i \in \Delta_4$ such that $\rho_\epsilon(\Gamma_{\hat{i}}) = 0$ (see [25] for details).

On the other hand, Theorem 5.2 easily proves that the class of PL 4-manifolds involved in Proposition 6.6(b) includes all PL 4-manifolds admitting semi-simple crystallizations.

It is an open problem to completely determine this second class of PL 4-manifolds for which additivity of regular genus holds.

References

[1] A. Akhmedov, B. Doug Park: *Exotic smooth structures on small* 4*-manifolds with odd signatures.* Invent. Math. **181**, (2010) 577–603.

[2] A. Akhmedov, M. Ishida, B. Doug Park: *Reducible smooth structures on* 4*-manifolds with zero signature.* J. Topology (2013). DOI 10.1112/jtopol/jtt043; First published online: October 31, 2013.

[3] P. Bandieri, M. R. Casali, C. Gagliardi: *Representing manifolds by crystallization theory: foundations, improvements and related results.* Atti Sem. Mat. Fis. Univ. Modena **49** (2001) 283–337.

[4] P. Bandieri, P. Cristofori, C. Gagliardi: *Nonorientable 3-manifolds admitting coloured triangulations with at most 30 tetrahedra.* J. Knot Theory Ramifications **18** (2009) 381–395.

[5] P. Bandieri, C. Gagliardi: *Rigid gems in dimension n.* Bol. Soc. Mat. Mexicana **18** (3) (2012) 55–67.

[6] B. Basak, M. R. Casali: *Lower bounds for regular genus and gem-complexity of PL* 4*-manifolds.* Preprint (2015), arXiv:1504.00771.

[7] B. Basak, J. Spreer: *Simple crystallizations of* 4*-manifolds.* Advances in Geometry (to appear), arXiv:1407.0752.

[8] B. A. Burton, J. Spreer: *Computationally proving triangulated* 4*-manifolds to be diffeomorphic.* arXiv:403.2780 [math.GT], 2013; 29th ACM Symposium on Computational Geometry, Young Researchers Forum, Collections of abstracts, 2013, 15–16.

[9] B. Benedetti, F. H. Lutz: *Random Discrete Morse Theory and a New Library of Triangulations.* Experimental Math. **23** (1) (2014) 66–94.

[10] S. E. Capell, J. L. Shaneson: *Some new four-manifolds.* Ann. of Math. (2) **104** (1976) 61-72.

[11] M. R. Casali: *A combinatorial characterization of 4-dimensional handlebodies.* Forum Math. **4** (1992) 123–134.

[12] M. R. Casali: *An infinite class of bounded* 4*-manifolds having regular genus three.* Bollettino Un. Mat. Ital. **10-A** (1996) 279–303.

[13] M. R. Casali: *Classifying PL 5-manifolds by regular genus: the boundary case.* Canadian J. Math. **49** (1997) 193–211.

[14] M. R. Casali: *Classification of non-orientable 3-manifolds admitting decompositions into* ≤ 26 *coloured tetrahedra.* Acta Appl. Math. **54** (1999) 75–97.

[15] M. R. Casali: *Dotted links, Heegaard diagrams and coloured graphs for PL* 4*-manifolds.* Revista Matematica Complutense **17** (2) (2004) 435–457.

[16] M. R. Casali: *Computing Matveev's complexity of non-orientable 3-manifolds via crystallization theory.* Topology Appl. **144** (2004) 201–209.

[17] M. R. Casali, P. Cristofori: *Computing Matveev's complexity via crystallization theory: the orientable case.* Acta Appl. Math. **92** (2) (2006) 113–123.

[18] M. R. Casali, P. Cristofori: *A catalogue of orientable 3-manifolds triangulated by* 30 *coloured tetrahedra.* J. Knot Theory Ramifications **17** (2008) 1–23.

[19] M. R. Casali, P. Cristofori: *A note about complexity of lens spaces.* Forum Math. (2014). DOI: 10.1515/forum-2013-0185, published online February 19, 2014, arXiv:1309.5728.

[20] M. R. Casali, P. Cristofori: *Cataloguing PL* 4*-manifolds by gem-complexity.* Electron. J. Combin. **22** (4) (2015) #P4.25, 1–25, arXiv:1408.0378.

[21] M. R. Casali, P. Cristofori, C. Gagliardi: *PL 4-manifolds admitting simple crystallizations: framed links and regular genus.* J. Knot Theory Ramifications (to appear), arXiv:1410.3321v2.

[22] M. R. Casali, P. Cristofori, M. Mulazzani: *Complexity computation for compact 3-manifolds via crystallizations and Heegaard diagrams.* Topology Appl. **159 (13)** (2012) 3042–3048. DOI: 10.1016/j.topol.2012.05.016.

[23] M. R. Casali, C. Gagliardi: *Classifying PL 5-manifolds up to regular genus seven.* Proc. Amer. Math. Soc. **120** (1994) 275–283.

[24] M. R. Casali, C. Gagliardi: *A code for m-bipartite edge-coloured graphs.* Rend. Ist. Mat. Univ. Trieste **32** (Suppl. 1) (2001) 55–76.

[25] M. R. Casali, L. Malagoli: *Handle-decompositions of PL* 4*-manifolds.* Cahiers Topologie Geom. Differentielle Categ. **38** (1997) 141–160.

[26] M. Casella, W. Kuhnel: *A triangulated* $K3$ *surface with the minimum number of vertices.* Topology **40** (4) (2001) 753–772.

[27] A. Cavicchioli: *A combinatorial characterization of* $\mathbb{S}^3 \times \mathbb{S}^1$ *among closed* 4*-manifolds.* Proc. Amer. Math. Soc. **105** (1989), 1008-1014.

[28] A. Cavicchioli: *On the genus of smooth* 4*-manifolds.* Trans. Amer. Math. Soc. **331** (1) (1992) 203–214.

[29] A. Cavicchioli, M. Meschiari: *On classification of* 4*-manifolds according to genus.* Cahiers Topologie Geom. Differentielle Categoriques **34** (1993) 37–56.

[30] M. Ferri, C. Gagliardi: *Crystallization moves.* Pacific J. Math. **100** (1982) 85–103.

[31] M. Ferri, C. Gagliardi: *The only genus zero -manifold is* $\mathbb{S}^n$. Proc. Amer. Math. Soc. **85** (1982) 638–642.

[32] M. Ferri, C. Gagliardi, L. Grasselli: *A graph-theoretical representation of PL-manifolds. A survey on crystallizations.* Aequationes Math. **31** (1986) 121–141.

[33] M. H. Freedman, F. Quinn: *Topology of* 4*-manifolds.* Princeton Mathematical Series. **39**, Princeton University Press, 1990.

[34] M. Furuta: *Monopole equation and the* $\frac{11}{8}$ *conjecture.* Math. Res. Lett. **8** (2001) 279–291.

[35] C. Gagliardi: *Extending the concept of genus to dimension* n. Proc. Amer. Math. Soc. **81** (1981) 473–481.

[36] C. Gagliardi: *On the genus of the complex projective plane.* Aequationes Math. **37** (1989) 130–140.

[37] C. Gagliardi: *How to deduce the fundamental group of a closed* n*-manifold from a contracted triangulation.* J. Comb. Inf. System Sci. **4** (1979) 237–252.

[38] C. Gagliardi, L. Grasselli: *Representing products of polyhedra by products of edge-coloured graphs.* Journal of Graph Theory **17** (5) (1993) 549–579.

[39] M. Kreck: *Some closed* 4*-manifolds with exotic differentiable structure.* In *Algebraic Topology, Aarhus 1982,* Proc. conference held in Aarhus, Denmark, 1982 (I. Madsen and B. Oliver, eds.). Lecture Notes in Mathematics **1051**, 246–262, Springer-Verlag, Berlin, 1984.

[40] P. B. Kronheimer, T. S. Mrowka: *Recurrence relations and asymptotics for four-manifold invariants.* Bull. Amer. Math. Soc. **30**(2) (1994) 215–221.

[41] S. Lins: *Gems, computers and attractors for 3-manifolds.* Knots and Everything **5**, World Scientific, 1995.

[42] S. Lins, M. Mulazzani: *Blobs and flips on gems.* J. Knot Theory Ramifications **15** (8) (2006) 1001–1035.

[43] R. Mandelbaum: *Four-dimensional topology: an introduction.* Bull. Amer. Math. Soc. **2** (1980) 1–159.

[44] A. Marani, M. Rivi, P. Cristofori: *Generation of Catalogues of PL* n*-manifolds: Computational Aspects on HPC Systems.* 14th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Timisoara (2012).

[45] J. M. Montesinos Amilibia: *Heegaard diagrams for closed* 4*-manifolds.* In *Geometric topology,* Academic Press, New York (1977) 219–237.

[46] J. Spreer, W. Kuhnel: *Combinatorial properties of the* $K3$ *surface, Simplicial blowups and slicings.* Experimental Math. **20** (2) (2011) 201–216.

On certain classes of closed 3-manifolds with different geometric structures

Alberto CAVICCHIOLI and Fulvia SPAGGIARI *

Department of Physics, Informatics and Mathematics
University of Modena and Reggio E.
Via Campi 213/B, 41125 Modena, Italy
alberto.cavicchioli@unimore.it fulvia.spaggiari@unimore.it

This paper is dedicated to one of the leading topologists of any time and over all a great gentleman, Prof. José Maria Montesinos, for his 70th birthday.

ABSTRACT

In this note, we review some recent results concerning the topology and geometry of closed connected orientable 3–manifolds. The used techniques are based on various combinatorial representations of 3–manifolds, such as polyhedral schemes, Heegaard diagrams, branched coverings, and Dehn surgery.

2010 Mathematics Subject Classification: 57M12, 57M25, 57M50.
Key words: 3-manifolds, Polyhedral schemes, Heegaard diagrams, branched coverings, Dehn surgery.

1. Combinatorial representations of 3-manifolds

(1.1) *Polyhedral schemes.* It is well-known that any closed connected 3–manifold M can be represented by a finite collection of tetrahedra whose 2–faces are identified in pairs via simplicial isomorphisms [28]. Of course, it is possible to construct M by using a single polyhedral 3–ball P whose finitely many boundary faces are glued together in pairs. The interior of P becomes an open 3–ball whose boundary meets itself in the manifold M along an embedded 2–complex K which is a *spine* of M. The

*Work performed under the auspices of the "National Group for Algebraic and Geometric Structures, and their Applications" of the GNSAGA-INDAM within the CNR (National Research Council) of Italy and partially supported by the MIUR (Ministero dell'Istruzione, dell'Universitá e della Ricerca) of Italy within the project "Strutture Geometriche, Combinatoria e loro Applicazioni".

manifold M is *orientable* exactly when the paired faces of P are oppositely oriented in its boundary. Of course, not every pairing of oppositely oriented boundary faces of a polyhedral 3–ball P yields an orientable 3–manifold. The only troublesome points in the resulting quotient space $M = P/\sim$ are the 0–cells of $K = (\partial P)/\sim$ that arise from the vertices of P. They have small regular neighbourhoods that are cones over certain closed surfaces. A classical result, proved by Seifert [28], states that the quotient space M is a closed orientable 3–manifold if and only if the Euler characteristic of M vanishes. More generally, one can represent M by a 3–dimensional full polyhedron having a pairwise associated polygonal faces. For this, we define a 3–dimensional full polyhedron to be a closed 3–ball whose boundary has been divided into polygons so that the following conditions are satisfied: each polygon is at least a 2–gon; each point of the 3–ball boundary belongs to at least one polygon; two polygons are either disjoint or have certain common edges or vertices. In the full polyhedron we let the faces be associated pairwise and require that the associations are topological mappings which transform vertices into vertices and edges into edges. Due to the pairing of faces, certain polyhedral edges and vertices will become equivalent to ones another. A result of this construction is that the quotient manifold M is uniquely determined topologically, up to homeomorphisms, by the scheme of the polyhedron. Finally, a finite presentation of the fundamental group of M can be derived from its polyhedral scheme. The fundamental group of M is that of the two–dimensional complex (spine) which arises from the boundary of the polyhedron by means of the identification of the equivalent faces.

(1.2) *Heegaard diagrams*. A *Heegaard splitting* of a closed connected orientable 3–manifold M is a pair (V, W) of homeomorphic orientable compact cubes with handles such that $M = V \cup W$ and $V \cap W = \partial V = \partial W$. The closed connected orientable surface $F = \partial V = \partial W$ is called the *Heegaard surface* of the splitting (V, W) of M. A classical theorem of Heegaard states that every closed connected orientable 3–manifold M admits a Heegaard splitting (see, for example, [27] and Chapter 5 of [9]). This splitting can be constructed as follows. Consider the 1–skeleton of a simplicial triangulation of M and define V as a regular neighbourhood of it. Then we set W to be the closure of the complement of V in M. The *Heegaard genus* $g(M)$ of M is the smallest integer g such that M has a Heegaard surface of genus g. Given a splitting (V, W) of M, let $D_1, \ldots, D_g$ be a collection of pairwise disjoint properly embedded discs in W which cut W into a 3–cell. The pairwise disjoint simple closed curves $\mathbf{w}_i = \partial D_i$ cut $F = \partial W$ into a 2–sphere with $2g$ holes. We say that $\mathbf{w} = \{\mathbf{w}_1, \ldots, \mathbf{w}_g\}$ is a set of *meridians* of the handlebody W. Let $\mathbf{v} = \{\mathbf{v}_1, \ldots, \mathbf{v}_g\}$ be a set of meridians of the handlebody V. Then the triple $(F, \mathbf{v}, \mathbf{w})$ is called a *Heegaard diagram* associated to the splitting (V, W) of M (or, briefly, a *Heegaard diagram* of M). The diagram can be drawn in a plane by flattening the above 2–sphere with $2g$ holes (whose quotient space is F). In this case, a set of meridians can be re–obtained by identifying in pairs the boundaries of the holes, while the other one gives rise to a set of pairwise disjoint simple arcs connecting the boundaries of the holes. The construction produces a planar graph (together with a pairing of the holes) which completely *represents* the manifold M in the sense that M can be recovered from it (for details see, for example, [9], Chp.5). Of course, there exist many different Heegaard diagrams representing the same manifold. The equivalence problem was solved by Singer in [29]: two different Heegaard diagrams of the same 3–manifold are related by a finite sequence of certain elementary moves (and/or their inverses), called *Singer's moves*. The first move changes the orientation on a curve of the diagram $(F, \mathbf{v}, \mathbf{w})$ or shifts a curve by isotopy. The second move substitutes a curve $\mathbf{w}_i$ of $\mathbf{w}$ with a curve $\mathbf{w'}_i$ after a light shifting to make $\mathbf{w'}_i$ disjoint to $\mathbf{w}_k$, for $i \neq k$. The curve $\mathbf{w'}_i$ is obtained by a connected sum of the curves $\mathbf{w}_i$ and $\mathbf{w}_k$. This operation is defined similarly for the set of meridians $\mathbf{v}$. The last move adds a trivial (i.e., unknotted) handle and a trivial curve to the diagram. It follows that Heegaard diagrams (up to Singer's moves) give an adequate representation of the closed connected orientable 3–manifolds in the sense that all invariants of the represented manifolds can be obtained from their diagrams via graph–theoretical algorithms. An interesting open problem in the theory of Heegaard diagrams is the famous conjecture that the Heegaard genus of a closed connected hyperbolic 3–manifold is equal to the number of minimal generators for the fundamental group.

(1.3) *Branched coverings*. Let M^n and N^n be compact manifolds with proper submanifolds $A^{n-2} \subset M$ and $B^{n-2} \subset N$. A continuous function $f : M \to N$ is said to be a *branched covering with branch sets A and B* if: (1) components of preimages of open sets of N are a basis for the topology of M; (2) $f(A) = B, f(M \backslash A) = N \backslash B$ and $N \backslash B$ is exactly the set of points in N which are evenly covered, i.e., have neighbourhoods U such that f sends each component of $f^{-1}(U)$ homeomorphically onto U. The restriction $f : M \backslash A \to N \backslash B$ is clearly a covering space, called the *associated unbranched covering*. By compactness of M, it is finite–sheeted. Each branch point $a \in A$ has a *branching index k*, meaning that f is k–to–one near a, and this number is constant on the components of A. Two branched covering spaces (M_i, f_i, N), $f_i : M_i \to N$, $i = 1, 2$, over N are said to be *equivalent* if there exists a homeomorphism $h : M_1 \to M_2$ with $f_2 \circ h = f_1$. A representation σ of a group G into the symmetric group S_k of order k is said to be *transitive* if the image $\sigma(G)$ acts transitively on the set of k letters used to define S_k. If M is a connected compact n–manifold and B a locally flat proper $(n-2)$–submanifold of M, then the equivalence classes of k–fold connected branched covering manifolds over M with branch set B correspond bijectively to the conjugacy classes of transitive representations of $G = \Pi_1(M \setminus B, x_0)$ into S_k sending each meridian of B in M to a non–trivial element of S_k. A transitive representation $\sigma : G \to S_k$ is called the *monodromy representation* of the associated branched covering space. Alexander proved that every closed oriented n–manifold M is a branched covering of the n–sphere (see, for example, [27]). The branching set is a $(n-2)$–subcomplex. Birman and Hilden [3] proved that for $n = 3$ the branching set can be assumed to be a closed submanifold that is a link in $\mathbb{S}^3$. A well–known theorem proved independently by Hilden [12] and Montesinos [20] states that every closed connected orientable 3–manifold M^3 is a 3–fold branched covering space over $\mathbb{S}^3$ with branch set a knot. It was proved in [22] that the Freudenthal compactification of an open connected oriented 3–manifold is a 3–fold branched covering of $\mathbb{S}^3$, and in some cases, a 2–fold branched covering of $\mathbb{S}^3$. The branching set is a locally finite disjoint union of strings. A set of covering moves that relate any two colored diagrams representing the same 3–manifold as simple branched 3–covering of $\mathbb{S}^3$ was described by Piergallini [23]. Connections between Heegaard diagrams and branched coverings can be found for example in [3], [18], [21], [27], and [30].

(1.4) *Dehn surgery.* Let $V = D^2 \times \mathbb{S}^1 \subset \mathbb{R}^3$ be the standard solid torus with meridian μ and longitude λ. A *meridian* of V is a simple closed curve on $\partial V \cong \mathbb{S}^1 \times \mathbb{S}^1$, non contractible in ∂V but homotopic to zero in V. A *longitude* of V is a simple closed curve on ∂V which intersects the meridian μ exactly in one point. We consider an oriented link $L = L_1 \cup L_2 \cup \cdots \cup L_n$ in the 3–sphere $\mathbb{S}^3 = \mathbb{R}^3 \cup \{\infty\}$, n disjoint tubular neighbourhoods $T_i \cong D^2 \times \mathbb{S}^1$ of L_i in $\mathbb{S}^3$, n copies V_i of V with meridian μ_i and longitude λ_i and n homeomorphisms $h_i : \partial V_i \to \partial T_i$ such that $h_{i\star}(\mu_i) = q_i \mathbf{m_i} + p_i \boldsymbol{\ell_i}$ where p_i and q_i are coprime integers and $(\mathbf{m_i}, \boldsymbol{\ell_i})$ is a preferred frame of T_i. A *preferred frame* $(\mathbf{m_i}, \boldsymbol{\ell_i})$ of T_i satisfies the following conditions: $\boldsymbol{\ell_i} \sim 0$ in $\mathbb{S}^3 \setminus L_i$ and the linking number between $\mathbf{m_i}$ and $\boldsymbol{\ell_i}$ equals 1. The rational number $r_i = q_i/p_i$ is called the *surgery coefficient associated to the component* L_i. The closed connected orientable 3–manifold

$$M = \left(\mathbb{S}^3 \setminus \bigcup_{i=1}^{n} int(T_i) \right) \bigcup_h \left(\bigcup_{i=1}^{n} V_i \right)$$

is the *result of a Dehn surgery along the link* $L = L_1 \cup L_2 \cup \cdots \cup L_n \subset \mathbb{S}^3$ with surgery coefficients $r_1 = q_1/p_1, r_2 = q_2/p_2, \ldots, r_n = q_n/p_n$. A well-known result states that every closed connected orientable 3–manifold can be obtained by Dehn surgery with surgery coefficients ± 1 along a link $L \subset \mathbb{S}^3$ whose components are unknotted. This result, known as the Lickorish-Wallace theorem, was first proved by Wallace [31] in 1960 and independently by Lickorish [15] in a stronger form in 1962 via the well-known relation between genuine surgery and cobordism. The Lickorish-Wallace theorem is equivalent to the fact that the oriented cobordism group of 3–manifolds is trivial, originally proved by Rokhlin in 1951. Two surgery descriptions are *equivalent* if one of them can be obtained from the other by a finite number of the following elementary moves: (1) add or delete a component with coefficient ∞; (2) do a twist on an unknotted component and change the surgery coefficients according to the Kirby–Rolfsen calculus (see [14] and [27]). Two surgery descriptions represent homeomorphic manifolds if and only if they are equivalent [14]. Connections between surgery on links and double branched coverings of $\mathbb{S}^3$ can be found in Montesinos [19].

2. Some classes of 3-manifolds with different geometries

(2.1) *Platonic Manifolds.* Let $X = \mathbb{S}^3$, $\mathbb{E}^3$ or $\mathbb{H}^3$. A *3–space form* (or, an *X–manifold*) M is an orbit space X/G, where G is an isometry group acting on X properly discontinuously and without fixed points. This gives a tiling of X. The isometries of G, mapping a distinguished tile onto its neighbours, identify the boundary faces of the tile in pairs. Such isometries generate the fundamental group of M. Of course, M is the quotient space obtained from any distinguished tile via the above pairing of its boundary faces. A *Platonic solid* in X is a polytope P with the combinatorial type of a Platonic solid (convex regular solid), embedded in X, so that all side lengths are equal, as are the interior face angles and dihedral angles. Everitt classified in [8], up to isometry, the orientable 3–space forms that arise from tilings of X by Platonic solids. This completes the work begun by several authors (see [2], [16], [25], and [26]). These results, obtained by algebraic and computational methods, follow from the classification of certain subgroups of rank four Coxeter groups. The following theorem summarizes the results of the quoted papers according to Everitt notation (explained after the statement of Theorem 2.3 below):

Theorem 2.1 (Everitt [8] Lorimer [16]) *The closed orientable spherical 3–manifolds arising from Platonic solids as space forms are listed in the following table:*

Spherical Manifolds	F	E	Homology
M_1	abab	a(- -)b(- -)aabb	$\mathbb{Z}_5$
M_2	ababcc	a(++)b(+-)aac(+-)bcd(+-)bcdd	$\mathbb{Z}_8$
M_3	abcbca	a(++)b(- -)c(+-)cd(- -)bdabdac	$\mathbb{Z}_2 \oplus \mathbb{Z}_2$
M_4	abcacbdd	a(++)b(+-)c(+-)ad(++)cbdacdb	$\mathbb{Z}_2 \oplus \mathbb{Z}_6$
M_5	abcacdbd	a(++)b(- -)c(++)ad(-+)cbcaddb	$\mathbb{Z}_8$
M_6	abcdcdab	a(++)b(++)c(++)d(++)bcdadabc	$\mathbb{Z}_3$
M_7	abcdefefbcda	a(-+)b(-+)c(-+)d(-+)e(-+)f(-+)g(-+) h(-+)i(-+)j(-+)idjefagbhcghijfeabcd	0
M_8	abcdefbdcfea	a(-+)b(-+)c(-+)d(-+)e(-+)f(++)g(++) h(++)i(++)j(++)ajcgbfeidhfhgjieabcd	$\mathbb{Z}_{15}$

The manifold M_1 comes from the tetrahedron with dihedral angle $2\pi/3$, M_2 and M_3 from the cube with angle $2\pi/3$, M_4, M_5 and M_6 from the octahedron with angle $2\pi/3$, and M_7 and M_8 from the dodecahedron with angle $2\pi/3$.

Remark 1. The manifolds M_7 and M_8 were constructed by Lorimer in [16]. The manifold M_3 is the quaternionic space [21], p.120, and M_6 is the octahedral space [21], p.117. The manifolds M_2 and M_5 have the same homology but they are not homeomorphic (hence non-isometric). This is proved in [8] by algebraic arguments which imply that $\pi_1(M_2) \cong \mathbb{Z}_8$ while $\pi_1(M_5)$ has order 24. We obtain all these facts as particular consequences of our geometric methods.

Theorem 2.2 (Everitt [8] Prok [25]) *The closed orientable Euclidean 3–manifolds arising from Platonic solids as space forms are listed in the following table:*

Euclidean Manifolds	F	E	Homology
M_9	abacbc	a(+++)b(+++)aac(+++)bccbcba	$\mathbb{Z}_3 \oplus \mathbb{Z}$
M_{10}	abbcca	a(-+-)ab(- -+)c(-+-)bacbbacc	$\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}$
M_{11}	abccba	a(-+-)ab(- -+)c(+- -)bccbbcaa	$\mathbb{Z}_4 \oplus \mathbb{Z}_4$
M_{12}	abcbca	a(+++)b(+++)c(+++)bcaaccbba	$\mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$
M_{13}	abcbca	a(+++)b(+++)c(-+-)cbaacbbca	$\mathbb{Z}_2 \oplus \mathbb{Z}$
$M_{14}(= M_{10})$	abcbca	a(-+-)b(+- -)c(+++)bcaaccbba	$\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}$

All these manifolds arise from the familiar cube with dihedral angle $2\pi/4$.

Remark 2. The manifold M_{12} is the 3–torus $\mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1$. The methods in [8] are not able to distinguish between the manifolds M_{10} and M_{14} (see [8], pp.260/261). Prok [25] constructed an Euclidean similarity between M_{10} and M_{14}, so they are indeed the same manifold. We prove again this fact in a different way.

Theorem 2.3 (Richardson and Rubinstein [26]) *The closed orientable hyperbolic 3–manifolds arising from Platonic solids as space forms are listed in the following table:*

Hyperbolic Manifolds	F	E	Homology
M_{15}	abcdefefbcda	a(-+-+)b(-+-+)c(-+-+)d(-+-+)e(-+-+) cdeabf(++++)afbfcfdfecdeabdeabc	$\mathbb{Z}_5 \oplus \mathbb{Z}_5 \oplus \mathbb{Z}_5$
M_{16}	abcdefdefbca	a(++++)b(++++)c(++++)d(++++)e(++++) abcdebf(++++)cfdfefafcdeabbcdea	$\mathbb{Z}_5 \oplus \mathbb{Z}_5 \oplus \mathbb{Z}_5$
M_{17}	abcdefdefbca	a(+-++)b(-+++)c(- - -+)d(++-+)e(+-++) debaf(+-++)bcfafefcdcfedabeabcd	$\mathbb{Z}_3 \oplus \mathbb{Z}_3$
M_{18}	abccadeefbfd	a(++- -)ab(-+++)ac(-+-+)d(-+++)bab e(+++-)ef(- -+-)bfdcaecdfffddcbece	$\mathbb{Z}_{35}$

M_{19}	abcdefebfdca	a(-+-+)b(-+-+)c(-+-+)d(-+-+)e(-+-+) edacbf(++++)cfefbfafdbdaeceabcd	$Z_5 \oplus Z_{15}$
M_{20}	abcdeffbdeca	a(++++)b(++++)c(++++)d(++++)e(++++) adbcecf(++++)efdfbfafeacdbdeabc	$Z_{15} \oplus Z_{15}$
M_{21}	abcdebedffca	a(+-++)b(+-++)c(- - -+)d(-+-+)e(-+++) cedaef(- -+-)afdfbfcfebdcbacdeab	$\mathbb{Z}_{48}$
M_{22}	abbcadefecfd	a(+++-)b(++-+)c(- -++)ad(-++-)a e(+-+-)dbbeaecf(+- -+)acfceffdedbdbfc	$\mathbb{Z}_{29}$
M_{23}	abcbdaefghihdefjgcji	a(-+)b(+-)c(- -)d(- -)e(-+)deabf(- +) g(+-)h(-+)i(+-)iaccj(++)jhdebfgfghij	$\mathbb{Z}_{11} \oplus \mathbb{Z}_{11}$
M_{24}	abcdebfceghhiijjfgda	a(-+)b(-+)c(-+)d(- -)e(++)cf(- -)ea g(- -)ebh(+ -)gi(+ -)dj(+ -)fghhdiifjjabc	$\mathbb{Z}_9$
M_{25}	abcdefbdgehiijjhfgca	a(++)b(++)c(++)d(++)e(+-)cdf(+-)ad g(+-)bfh(-+)gi(+-)ej(-+)ijgjhehifabc	$\mathbb{Z}_2 \oplus \mathbb{Z}_{18}$
M_{26}	abcdaefdgfhihcjjbige	a(++)b(+-)bc(+-)d(- -)e(+-)baf(- -) g(+-)efgh(++)ghci(+-)dj(-+)jjdeiicahf	$\mathbb{Z}_{35}$
M_{27}	abcdabefghcijidfjghe	a(++)ab(-+)c(++)d(++)e(- -)bacf(+-) g(+-)h(+-)ei(-+)j(++)djfidhgihebgjfc	$\mathbb{Z}_{29}$
M_{28}	abcdaebdfghicjehjfgi	a(++)b(+-)bc(- -)d(-+)e(++)bacdef(+-) g(- -)h(+-)di(-+)aj(- -)ijfehgcighjf	$\mathbb{Z}_{29}$

The manifolds $M_{15}, \dots, M_{22}$ come from the dodecahedron with dihedral angle $2\pi/5$, and $M_{23}, \dots, M_{28}$ from the icosahedron with angle $2\pi/3$.

Remark 3. For manifolds in Theorem 2.3 with the same first homology, algebraic arguments are provided in [26] to show that they are all distinct. The manifold M_{15} is the Seifert-Weber 3–manifold [10].

Now we recall the Everitt notation for the tables. The columns F and E give the face and edge identifications in the form of an encoded string of letters and $\pm$ signs to be read in conjunction with Figure 1. The ith and jth faces are paired when the ith and jth positions of the string in column F are occupied by the same letter. Similarly, for the edge identifications, where a string of $\pm$´ s after a letter indicates whether the corresponding edge is identified with subsequent ones with the orientations matching or reversed. For example, the Seifert-Weber manifold M_{15} arising from the dodecahedron with dihedral angle $2\pi/5$ has face identifications $abcdefefbcda$, where a indicates that faces 1, 12 are identified, b indicates that faces 2, 9 are identified, and so on. It has edge identifications

$$\mathrm{a}(-+-+)\mathrm{b}(-+-+)\mathrm{c}(-+-+)\mathrm{d}(-+-+)\mathrm{e}(-+-+)$$

$$\mathrm{cdeabf}(++++)\mathrm{afbfcfdfecdeabdeabc}$$

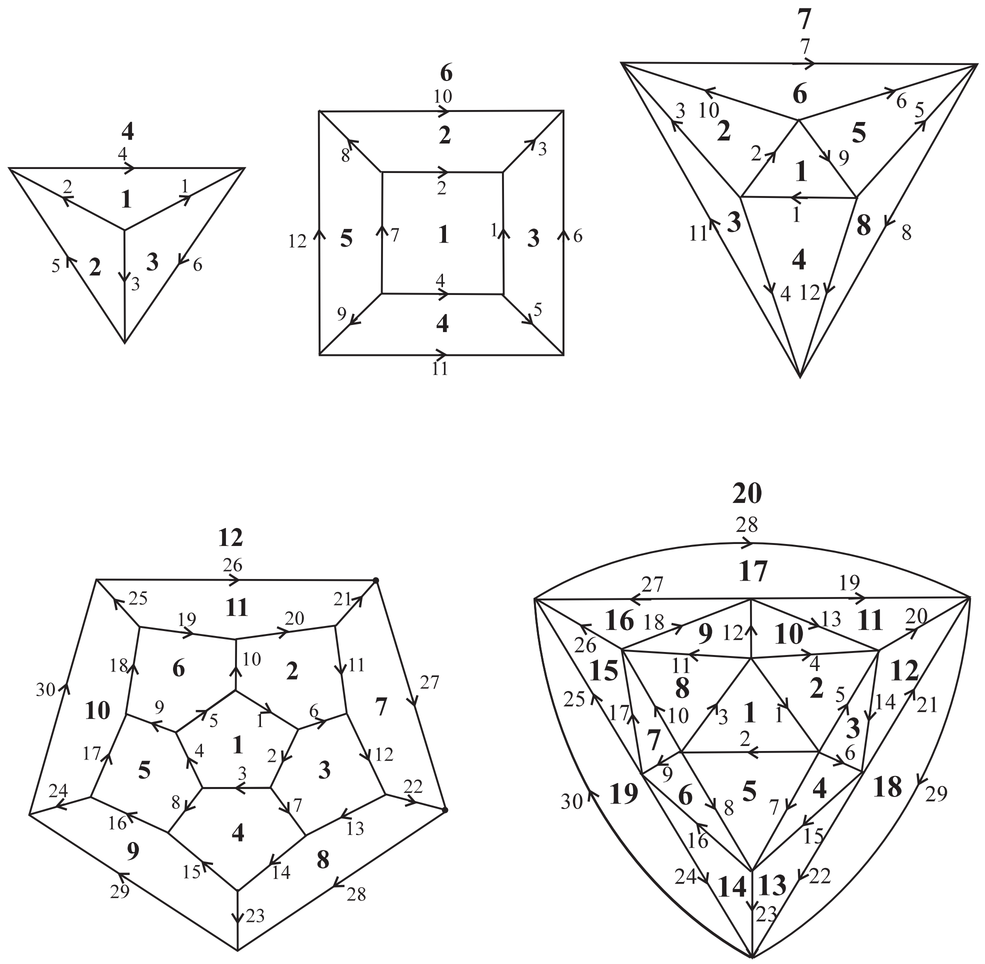


Figure 1. Platonic solids labelled according to Everitt notation

where a indicates that edges 1, 9, 12, 24 and 28 are identified, and $a(-+-+)$ means edge 1 is identified with edge 9 so that the identifications are reversed, with edge 12 so they match, with edge 24 so they are reversed, and with edge 28 so they match. From the data in these two columns one can reconstruct the side pairing of the boundary faces of the Platonic solid. This completely defines the quotient manifold. From the polyhedral representation, one obtain immediately a finite presentation of the fundamental group and a Heegaard diagram of the quotient manifold.

Theorem 2.4 [6] *The spherical and Euclidean manifolds obtained from Platonic solids as space forms are homeomorphic to the following fibered spaces:*

Spherical manifolds
$M_1 \cong L(5,2)$
$M_2 \cong L(8,3)$
$M_3 \cong \mathbb{S}^3/_{\langle 222 \rangle} = (O\ 0\ o\ : -1\ (2,1)(2,1)(2,1))$
$M_4 \cong \mathbb{S}^3/_{Q_8 \times \mathbb{Z}_3} = (O\ 0\ o\ : 0\ (2,1)(2,1)(2,1))$
$M_5 \cong \mathbb{S}^3/_{D_{24}} = (O\ 0\ o\ : -1\ (2,1)(2,1)(3,2))$
$M_6 \cong \mathbb{S}^3/_{\langle 332 \rangle} = (O\ 0\ o\ : -1\ (3,1)(3,1)(2,1))$
$M_7 \cong \mathbb{S}^3/_{P_{120}} = (O\ 0\ o\ : -1\ (2,1)(3,1)(5,1))$
$M_8 \cong \mathbb{S}^3/_{P_{24} \times \mathbb{Z}_5} = (O\ 0\ o\ : -1\ (2,1)(3,2)(3,2))$

Euclidean manifolds
$M_9 \cong T \times I/_{\left(\begin{smallmatrix} 0 & 1 \\ -1 & -1 \end{smallmatrix}\right)} = (O\ 0\ o\ : -1\ (3,1)(3,1)(3,1))$
$M_{10} \cong M_{14} \cong T \times I/_{\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)} = (O\ 0\ o\ : -2\ (2,1)(2,1)(2,1)(2,1))$
$M_{11} \cong (K \underset{\sim}{\times} I) \cup (K \underset{\sim}{\times} I)/_{\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)} = (O\ 1\ n\ : -1\ (2,1)(2,1))$
$M_{12} \cong T \times I/_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)} = \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1$
$M_{13} \cong T \times I/_{\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)} = (O\ 0\ o\ : -1\ (4,1)(4,1)(2,1))$

Theorem 2.5 [6] *For the closed orientable hyperbolic 3–manifolds arising from Platonic solids as space forms, the following properties hold:*

a) *The manifolds M_{15} (Seifert-Weber) and M_{16} coincide with the manifolds $M_{5,2}$ and $M_{5,1}$, respectively, constructed in* [10]. *They are 5–fold strongly cyclic coverings of the 3–sphere branched over the Whitehead link. These manifolds have the same homology but they are distinct;*

b) *The manifold M_{20} is the Lorimer dodecahedral space with homology $\mathbb{Z}_{15} \oplus \mathbb{Z}_{15}$* ([16])*;*

c) *The manifold M_{23} is the Fibonacci manifold M_5 (of Heegaard genus 2) encoded by the standard presentation of the Fibonacci group $F(2,10)$ with generators $x_1, \dots, x_{10}$ and relations $x_i x_{i+1} = x_{i+2}$ (subscripts mod 10). It is the 5–fold (resp. 2–fold) cyclic covering of the 3–sphere branched over the figure eight knot (resp. the knot 10_{123})(see* [11] *and* [13]*);*

d) *The manifolds M_{24} and M_{25} are 3–fold strongly cyclic coverings of the lens space $L(3,1)$ branched over two (non-equivalent) 2–component links;*

e) *The manifolds M_{26}, M_{27}, and M_{28} have Heegaard genus 2, and they are 2–fold coverings of the 3–sphere branched over the π–hyperbolic 3–bridge knots K_{26}, K_{27}, and K_{28}, respectively, depicted in Figure 2. The knots K_{26} and K_{28} are chiral and invertible while K_{27} is chiral and non-invertible. The symmetry group of K_{27} and K_{28} (resp. K_{26}) is $\mathbb{Z}_2$ (resp. D_2). The manifolds M_{27} and M_{28} have the same homology but they are distinct.*

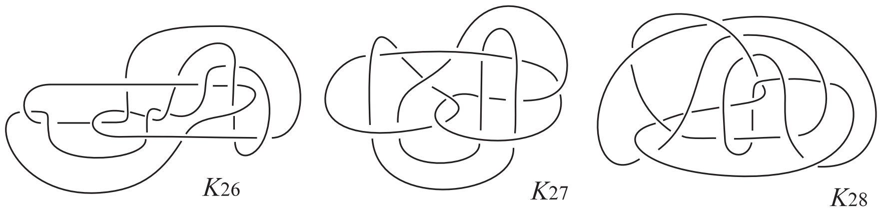


Figure 2.

(2.2) *Hyperbolic tetrahedron manifolds.* The results in this subsection were obtained in [5]. For positive integers m and n, we consider a simplicial complex $P_{m,n}$ which triangulates the standard tetrahedron, see Figure 3. Let A_0, A_1, A_2 and A_3 be the vertices of the tetrahedron. The oriented edge A_0A_1 (resp. A_2A_3) is subdivided into $2m+1$ edges, alternatively labelled by x_1 and x_2, (resp. x_2 and x_1), the oriented edges A_1A_2 and A_0A_3 are subdivided into $2n+2$ edges, alternatively labelled by x_2 and x_1, and the oriented edges A_1A_3 and A_0A_2 are subdivided into $2m$ edges, alternatively labelled by x_1 and x_2. We identify in pairs the faces of the boundary of the tetrahedron via the orientation reversing homeomorphisms a and b. The faces are to be paired so that the index stars in Figure 3 match up. The resulting identification space, denoted by $M_{m,n}$, has one vertex, two 1–cells, also denoted by x_1 and x_2, two 2–cells, and one 3–cell. Since the Euler characteristic vanishes, we get a closed connected orientable 3–manifold $M_{m,n}$ by a well-known criterion obtained in [28]. See also Subsection (1.1). The face pairing forces an identification of the cells of $P_{m,n}$ into a 2–polyhedron which is a spine of the obtained manifold. We can obtain a finite presentation for the fundamental group $G_{m,n}$ of $M_{m,n}$ by considering the face pairing generators a and b.

Theorem 2.6 *For $m,n \geq 1$, the simplicial complex $P_{m,n}$ with the identifications described above defines the tetrahedron manifold $M_{m,n}$ which has a spine modeled on the finite presentation*

$$G_{m,n} = \langle a,\ b, :\ (ab)^n ab^2(a^{-2}b^2)^m = 1,\ (ba)^n ba^2(b^{-2}a^2)^m = 1\rangle.$$

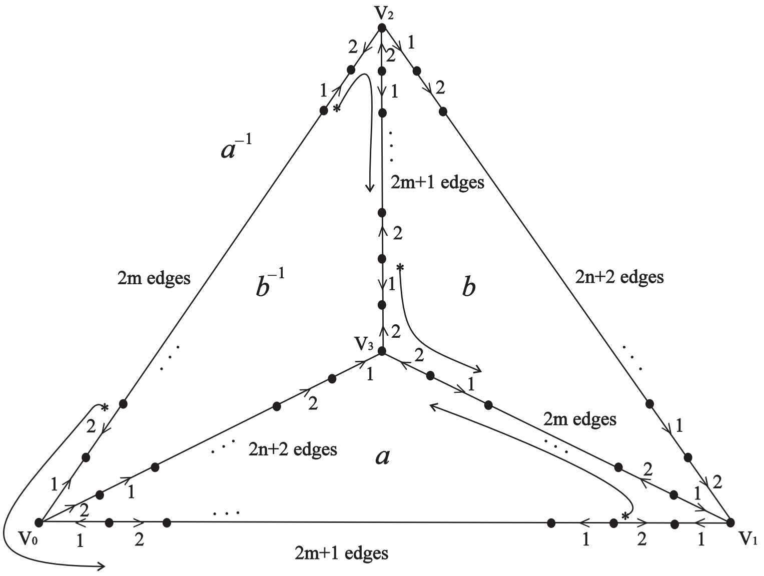


Figure 3. Polyhedral schemata of the tetrahedron manifolds $M_{m,n}$

The following result states that the constructed tetrahedron manifolds can be represented as 2–fold branched covering of the 3–sphere.

Theorem 2.7 *For $m, n \geq 1$, the tetrahedron manifold $M_{m,n}$ is the 2–fold covering of the 3–sphere branched over the hyperbolic knot $K_{m,n}$ depicted in Figure 4.*

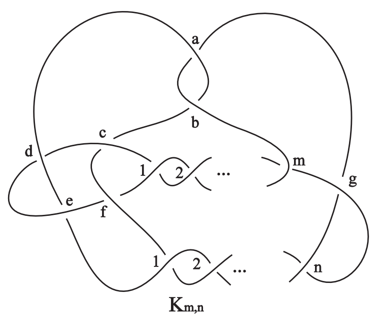


Figure 4. The knot $K_{m,n}$

To give a surgery description of our tetrahedron manifolds $M_{m,n}$ for $m, n \geq 1$, we use a well-known construction due to Montesinos [19]

Theorem 2.8 *For $m, n \geq 1$, the closed tetrahedron manifold $M_{m,n}$ can be obtained by $(2n+3)/(n+1)$ and $(4m+1)/m$ Dehn surgeries on the components of the Whitehead link $\mathcal{W}$.*

(2.3) *Tetrahedron manifolds with Sol geometry.* The results of this subsection were obtained in [4]. For every $n \in \mathbb{N}$, $n > 2$, let us consider the simplicial complex P_n which triangulates the boundary of the standard tetrahedron as indicated in Figure 5. Let A_0, A_1, A_2 and A_3 be the vertices of the tetrahedron. The oriented edges A_0A_3 and A_1A_2 are subdivided into three edges labeled by the oriented sequences $a^{-1}ba$ and aba^{-1}, respectively. The oriented edges A_0A_1 and A_2A_3 are subdivided into four edges labeled by the oriented sequences $b^{-1}a^{-1}ba$ and $bab^{-1}a^{-1}$, respectively. The oriented edge A_1A_3 is labeled by b, and the oriented edge A_0A_2 is subdivided into $n-1$ edges labeled by the oriented sequence b^{n-1}. We identify in pairs the boundary faces of the tetrahedron via the orientation-reversing homeomorphisms $x : x^{-1} \to x$ and $y : y^{-1} \to y$. The faces are to be paired so that the index stars in Figure 5 match up. The resulting space, denoted by $\overline{M}_n$, has one vertex, two 1–cells, also denoted by a and b, two 2–cells, and one 3–cell. Since the Euler characteristic vanishes, $\overline{M}_n$ is a closed connected (orientable) 3–manifold. The face pairing forces an identification of the cells of P_n into a 2–polyhedron which is a spine of $\overline{M}_n$. We can immediately obtain a finite presentation G_n for $\pi_1(\overline{M}_n)$ by considering the face pairing generators x and y by two so–called Poincaré cycles in the universal covering space or, equivalently, by two spine relations above.

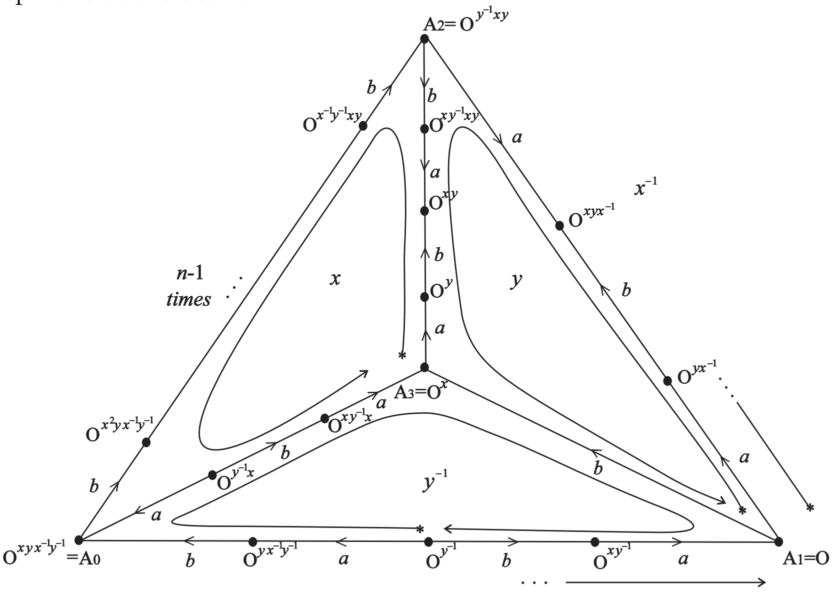


Figure 5. Polyhedral schemata of the tetrahedron manifolds $\overline{M}_n$

Theorem 2.9 *For any natural number $n > 2$, the simplicial complex P_n with the identifications described above defines the tetrahedron manifold $\overline{M}_n$ with two cycle relations around edges a and b, respectively, providing the finite presentation*

$$G_n = \langle x,\ y\colon\ xyx^{-1}y^{-1}x^{-1}yxy^{-1} = 1, \quad yxy^{-2}xyx^{-n} = 1\rangle.$$

We showed in [4] that the above presentations are geometric, that is, they are induced by Heegaard diagrams (of genus 2) representing our manifolds. Starting from such diagrams and using standard geometric constructions, we derive the following covering property of the constructed tetrahedron manifolds:

Theorem 2.10 *For any $n > 2$, the tetrahedron manifold $\overline{M}_n$ is the 2–fold covering of the 3–sphere branched over the 3–bridge link L_n depicted in Figure 6. If n is odd (resp. even), then L_n has two (resp. three) components.*

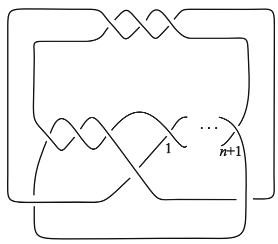


Figure 6. The link L_n, $n > 2$

The link L_n in Figure 6 is exactly the link pictured on Figure 7 from [17] for the case $p/q = 0$ and $m/n = n - 2$ (here n on the left side of the formula is from [17], but n on the right side is ours). So by Theorem 2 from [17] we get immediately a surgery description of our tetrahedron manifolds $\overline{M}_n$, $n > 2$.

Theorem 2.11 *The tetrahedron manifold $\overline{M}_n$, $n > 2$, can be obtained by $n - 2$ and 0 (zero) Dehn surgeries on the components of the Whitehead link $\mathcal{W}$.*

Let $\mathcal{O} = \mathcal{W}(0, n - 2)$ be the closed orbifold whose underlying space is the 3–sphere and the singular set is the Whitehead link with branching coefficients 0 and $n - 2$ on its components. It is well-known that the Whitehead link has a Montesinos fibration. This induces an orbifold Seifert fibration of $\mathcal{O}$. So the orbifold $\mathcal{O}$ and hence the manifold $\overline{M}_n$ admit a geometric structure (see, for example, [24], p.847).

Theorem 2.12 *The tetrahedron manifold $\overline{M}_n$, $n > 2$, is diffeomorphic to the mapping torus M_φ of the Anosov linear diffeomorphism of the torus $T \cong \mathbb{R}^2/\mathbb{Z}^2$ defined by the integer matrix $\begin{pmatrix} 0 & -1 \\ 1 & n \end{pmatrix}$. Then $\overline{M}_n$ admits a geometric structure modeled on real* **Sol** *geometry.*

(2.4) *Hyperbolic surgery manifolds.* The results in this subsection were obtained in [7]. For any positive integer n, let $\mathcal{L}_{2n+1}$ be the oriented link with $2n+1$ components L_0, L_i and K_i, $i = 1, \ldots, n$, in the oriented 3–sphere $\mathbb{S}^3$ depicted in Figure 7. This link can be obtained as a belted sum of Borromean rings, as remarked in [1], p.8. thus, it is hyperbolic for any $n \geq 1$. Let us consider the closed connected orientable

3–manifolds $M_n(r_i/s_i; p_i/q_i; h/k)$ obtained by Dehn surgery on $\mathbb{S}^3$ along the oriented link $\mathcal{L}_{2n+1}$ such that the surgery coefficients r_i/s_i , p_i/q_i and h/k correspond to the oriented components L_i, K_i and L_0, respectively, where $i = 1, \ldots, n$. Of course, we always assume that $gcd(r_i, s_i) = 1$, $gcd(p_i, q_i) = 1$ and $gcd(h, k) = 1$.

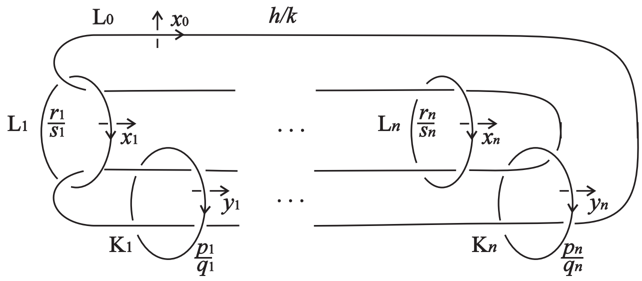


Figure 7. Dehn surgery description of the 3–manifold $M_n(r_i/s_i; p_i/q_i; h/k)$

This family of manifolds contains all closed manifolds obtained by Dehn surgeries with rational coefficients along the 2–bridge knots.

Theorem 2.13 *The fundamental group of the surgery 3–dimensional manifold* $M_n(r_i/s_i; p_i/q_i; h/k)$ *admits the finite balanced presentation with* $2n + 1$ *generators* a_i, b_i, *and* c, $i = 1, \ldots, xn$, *and* $2n + 1$ *relations:*

$$a_i^{r_i} b_i^{q_i} \cdots b_n^{q_n} c^k b_n^{-q_n} \cdots b_i^{-q_i} c^{-k} = 1$$

$$b_i^{p_i} c^{-k} a_i^{-s_i} \cdots a_1^{-s_1} c^k a_1^{s_1} \cdots a_i^{s_i} = 1$$

$$c^h b_n^{-q_n} \cdots b_1^{-q_1} a_1^{s_1} b_1^{q_1} \cdots a_n^{s_n} b_n^{q_n} a_n^{-s_n} \cdots a_1^{-s_1} = 1.$$

The closed manifold $M_n(r_i/s_i; p_i/q_i; h/k)$ *admits a Heegaard diagram of genus* $2n+1$ *inducing the above presentation, which is thus geometric. Furthermore, the Heegaard genus of* $M_n(r_i/s_i; p_i/q_i; h/k)$ *is at most* $2n + 1$.

As remarked in [1], p.8, the link $\mathcal{L}_{2n+1}$ is hyperbolic in the sense that it has a hyperbolic complement. So the Thurston–Jorgensen theory [30] of hyperbolic surgery gives the following result:

Theorem 2.14 *For any integer* $n \geq 1$, *and for almost all pairs of surgery coefficients* r_i/s_i, p_i/q_i *and* h/k, *the closed connected orientable* 3*–manifolds* $M_n(r_i/s_i; p_i/q_i; h/k)$ *are hyperbolic.*

The next results describe some covering properties of our surgery manifolds. Using Montesinos' trick [19], we prove that such manifolds are 2–fold branched covers of a connected sum of lens spaces. Moreover, it follows that a large subclass of our surgery manifolds are 2–fold coverings of the 3–sphere branched over well-specified clearly depicted links.

Theorem 2.15 *Suppose that r_i is odd for every $i = 1, \ldots, n$. Then the surgery manifold $M_n(r_i/s_i; p_i/q_i; h/k$ is 2–fold branched covering of the connected sum of n lens spaces $L(r_1, 2s_1)\# \ldots \# L(r_n, 2s_n)$.*

Theorem 2.16 *Let $\mathcal{M} = M_n(r_i/s_i; p_i/q_i; h/k)$, $r_i = 1$ and $s_i \geq 1$, for $i = 1, \ldots, n$, be the closed connected orientable 3–manifold obtained by Dehn surgeries on the components of the link $\mathcal{L}_{2n+1}$. Then $\mathcal{M}$ is the 2–fold covering of the 3–sphere branched over the link $\mathcal{L}_r(p_i/q_i; h/k)$, where $r = 2s_1 + \cdots + 2s_n$, pictured in Figure 8.*

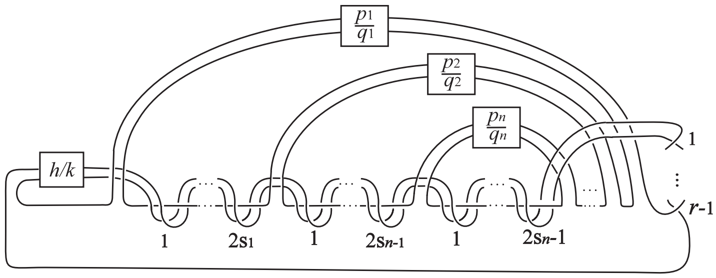


Figure 8. The link $\mathcal{L}_r(p_i/q_i; h/k)$, $r = 2s_1 + \cdots + 2s_n$

References

[1] C. Adams: *Hyperbolic Knots*. arXiv math/GT Geometric Topology, math/0309466 (2003).

[2] L. A. Best: *On torsion-free discrete subgroups of $PSL_2(\mathbb{C})$ with compact orbit space.* Canad. J. Math. **23** (1971) 451–460.

[3] J. S. Birman, H. M. Hilden: *Heegaard splittings of branched coverings of* $\mathbb{S}^3$. Trans. AMS **213** (1975) 315–352.

[4] A. Cavicchioli, E. Molnár, F. Spaggiari, J. Szirmai: *Some tetrahedron manifolds with Sol geometry and related groups.* Journal of Geometry **105** (2014) 601–614.

[5] A. Cavicchioli, F. Spaggiari: *Tetrahedron manifolds series of Heegaard genus two with knot presentation and Dehn surgery.* Acta Math. Hungarica **131(4)** (2011) 307–322.

[6] A. Cavicchioli, F. Spaggiari, A. I. Telloni: *Topology of compact space forms from platonic solids. I, II.* Topology Appl. **156** (2009) 812–822, **157** (2010) 921–931.

[7] A. Cavicchioli, F. Spaggiari, A. I. Telloni: *Fundamental group and covering properties of hyperbolic surgery manifolds.* Geometry **2013** (2013) 1–8.

[8] B. Everitt: *3-manifolds from Platonic solids.* Topology Appl. **138** (2004) 253–263.

[9] A. T. Fomenko, S. V. Matveev: *Algorithmic and Computer Methods for Three–Manifolds.* Math. and Appl. **425** (1997), Kluwer Acad. Publ., Dordrecht-Boston-London.

[10] H. Helling, A. C. Kim, J. L. Mennicke: *Some honey-combs in hyperbolic 3-space.* Commun. Algebra **23** (1995) 5169–5206.

[11] H. Helling, A. C. Kim, J. L. Mennicke: *A geometric study of Fibonacci groups*. J. Lie Theory **8** (1998) 1–23.

[12] H. M. Hilden: *Three-fold branched coverings of* $\mathbb{S}^3$. Amer. J. Math. **98** (1976) 989–997.

[13] H. M. Hilden, M. T. Lozano, J. M. Montesinos-Amilibia: *The arithmeticity of the figure eight knot orbifolds*. In: *Topology '90* (eds. B. Apanasov, W. D. Neumann, A. W. Reid, L. Siebenmann) Ohio State Univ., Math. Research Inst. Publ., Walter de Gruyter Ed., Berlin **1** (1992) 169–183.

[14] R. Kirby: *A calculus for framed links in* $\mathbb{S}^3$. Invent. Math. **45** (1978) 35–56.

[15] W. B. R. Lickorish: *A representation of orientable combinatorial 3–manifolds*. Ann. Math. **76**(2) (1962) 531–540.

[16] P. J. Lorimer: *Four dodecahedral spaces*. Pacific J. Math. **156** (2) (1992) 329–335.

[17] A. D. Mednykh, A. Yu. Vesnin: *Covering properties of small volume hyperbolic 3–manifolds*. J. Knot Theory and its Ram. **7**(3) (1998) 381–392.

[18] J. M. Montesinos: *Variedades des Seifert que son recubridores de dos hojas*. Bol. Soc. Mat. Mexicana **18** (1973) 1–32.

[19] J. M. Montesinos: *Surgery on links and double branched covers of* $\mathbb{S}^3$. In: *Knots, Groups and 3-Manifolds,* (L. P. Neuwirth, ed.), Ann. Math. Studies **84**, Princeton Univ. Press, Princeton, NJ, 1975.

[20] J. M. Montesinos: *Three-manifolds as 3-fold branched covers of* $\mathbb{S}^3$. Quart. J. Math. Oxford Ser.2 **27** (1976) 85–94.

[21] J. M. Montesinos: *Classical Tessellations and Three–Manifolds*. Universitext, Springer-Verlag, Berlin, 1987.

[22] J. M. Montesinos: *Representing open 3-manifolds as 3-fold branched coverings*. Revista Matematica Complutense **15**(2) (2002) 533–542.

[23] R. Piergallini: *Covering moves*. Trans. AMS **325** (2) (1991) 903–920.

[24] J. Porti: *Regenerating hyperbolic cone structures from Nil*. Geometry & Topology **6** (2002) 815-852.

[25] I. Prok: *Classification of dodecahedral space forms*. Beiträge Algebra Geom. **39** (2) (1998) 497–515.

[26] J. Richardson, J. H. Rubinstein: *Hyperbolic manifolds from a regular polyhedron*. Preprint.

[27] D. Rolfsen: *Knots and Links*. Math. Lect. Series **7**, Publish or Perish Inc., Berkeley, 1976.

[28] H. Seifert, W. Threlfall: *A Textbook of Topology; Topology of 3–dimensional fibered spaces*. Academic Press, New York-London, 1980.

[29] J. Singer: *Three-dimensional manifolds and their Heegaard diagrams*. Trans. Amer. Math. Soc. **35** (1933) 88–111.

[30] W. P. Thurston: *The Geometry and Topology of* 3*–Manifolds*. Lect. Notes, Princeton Univ. Press, Princeton, NJ, USA, 1980.

[31] A. H. Wallace: *Modifications and Cobounding Manifolds*. Canadian J. Math. **12** (1960) 503–552.

Algunas exploraciones matemáticas del mundo

Capi Corrales Rodrigáñez

Departamento de Álgebra, Facultad de Matemáticas
Universidad Complutense de Madrid
28040 Madrid
capi@mat.ucm.es

Para José María Montesinos Amilibia.

Fig. 1: Amadeo Olmos 2011

1. Introducción

> "La filosofía[1] está escrita en ese grandísimo libro que tenemos abierto ante los ojos, quiero decir, el universo, pero no se puede entender si antes no se aprende a entender la lengua, a conocer los caracteres en los que está escrito. Está escrito en lengua matemática y sus caracteres son triángulos, círculos y otras figuras geométricas, sin las cuales es imposible entender ni una palabra; sin ellos es como girar vanamente en un oscuro laberinto." (Galileo Galilei, *Il saggitore*, cap. 6-4.)

No hay más que entrar en el despacho de José María Montesinos, asistir a una de sus conferencias o leer alguno de sus trabajos científicos, para saber que es uno de esos exploradores que utilizan las matemáticas para cartografiar el mundo. Uno

[1] En época de Galileo, sinónimo de conocimiento.

de los grandes constructores de mapas del universo, fue el matemático Johannes Kepler (1571-1630), contemporáneo de Galileo y voraz lector del libro de la naturaleza. Kepler dedicó su vida a buscar estructuras e ideas tras las cosas que hiciesen posible utilizar las matemáticas para describirlas y/o entender por qué son como son.

> "En tiempos de ansiedad e incertidumbre como los nuestros, cuando tan difícil es sentir satisfacción por el curso de la humanidad, resulta particularmente consolador pensar en un hombre tan excepcional y sereno como Kepler. [...] Al parecer, la mente humana ha de construir primero las formas de modo independiente, para luego poder hallarlas en las cosas. Las verdaderas proezas de Kepler son un ejemplo magnífico de esta afirmación: el conocimiento no puede surgir de la experiencia tan sólo, sino de la comparación de las invenciones del intelecto con los hechos observados. (A. Einstein, *Johannes Kepler*, [3], p. 254.)

En 1610, Kepler escribió *Sobre el copo de nieve de seis picos. Un regalo de Año Nuevo* ([6]), un delicioso tratado tan corto en extensión como rico en ideas. Considerado hoy precursor de los estudios de glaceología y mineralogía cristalina, este breve tratado marca un hito en el uso de las matemáticas para entender el mundo físico que nos rodea. En él, Kepler describe el recorrido de preguntas y reflexiones que siguió en su estudio de los copos de nieve y, de paso, ilustra elegantemente cómo formular en lenguaje matemático nuestras reflexiones y preguntas sobre la naturaleza.

En el proceso de analizar lo que él denomina *la facultad formativa que actúa en el universo* –concretamente en la formación de los copos de nieve–, Kepler nos lleva a leer con nuevos ojos –matemáticos– textos como *El Timeo* de Platón, *Elementos* de Euclides, o el Libro V de *Colección* de Pappus. En matemáticas, las buenas soluciones a problemas se caracterizan por dar lugar, casi de inmediato, a nuevas preguntas. Las respuestas que Kepler encontró en estas obras clásicas a algunas de sus dudas, le llevaron a proponer(se) nuevos retos que describe –algunas veces incluyendo la solución– en *El copo de nieve de seis picos*. El descubrimiento de dos nuevos poliedros semirregulares, los primeros enunciados de las famosas Conjetura de Kepler y Conjetura del Panal y la primera manifestación escrita que conocemos de la relación entre la serie de Fibonacci y la proporción aúrea, son algunas de las joyas que podemos disfrutar en este tratado. Todas ellas, de una manera u otra, están relacionadas con los temas que José María Montesinos Amilibia encara en sus investigaciones. Es por esto que se me ha ocurrido utilizar el tratado de Kepler como punto de partida a partir del cual elaborar, como regalo a José María, un pequeño recorrido por algunos textos escritos por quiénes le precedieron en sus exploraciones geométricas del universo[2].

Agradezco a Amadeo Olmos (Segovia) que me haya permitido reproducir algunas de sus obras en este trabajo, a Fernando Porqueras (Gabinete Informático, Facultad de Matemáticas, UCM) su ayuda técnica y a José Manuel Gamboa (Departamento

[2]Muchas de las citas originales que aparecerán en estas páginas, las he traducido al castellano yo misma a partir las traducciones al inglés o francés recogidas en la bibiografía

de Álgebra, Facultad de Matemáticas, UCM) la revisión de la primera versión de este texto.

2. El recorrido

"Sé muy bien lo partidario que es usted de Nada... Por lo que puedo deducir fácilmente que tanto más grato le resultará un regalo cuanto más se acerque a Nada. Lo que sea, pues, que le agrade a usted por su evocación de Nada, habrá de ser pequeño e insignificante, barato y fugaz — es decir, casi Nada. Y puesto que hay muchas cosas así en el dominio de la naturaleza, hay que elegir una de ellas. Mientras consideraba estas cuestiones con ansiedad, crucé el puente, preocupado por mi falta de civismo al aparecer ante usted sin un regalo de Año Nuevo, excepto, quizás (y por seguir tocando la misma nota), el que siempre le llevo... o sea, Nada. Tampoco era capaz de pensar en algo que siendo cercano a Nada, diese lugar a sutil reflexión. Justo entonces, por feliz ocurrencia, parte del vapor del aire se acumuló en nieve por la fuerza del frío, y unos cuantos copos dispersos cayeron sobre mi abrigo, todos con seis picos y radios moñudos.

Fig. 2: Amadeo Olmos 2009

¡Por Hércules! Aquí había algo más pequeño que una gota, y sin embargo dotado de forma. Aquí, sin duda, estaba el más deseable regalo de Año Nuevo para el amante de Nada, y uno digno también de un matemático (que Nada tiene, y Nada recibe), pues desciende del cielo y se parece a las estrellas. Nuestra pregunta será por qué los copos de nieve, cuando acaban de caer y antes de amontonarse, siempre caen con seis picos y con seis radios moñudos como plumas.

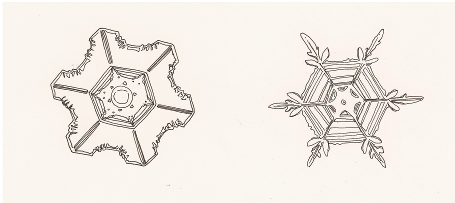


Fig. 3

Utilizaremos algunos ejemplos bien conocidos y los presentaremos de una manera geométrica, pues hacerlo beneficiará enormemente nuestra investigación.

[...] Es manifiesto que las semillas de la granada son presionadas en forma rómbica por la conjunción de necesidad material y el principio de crecimiento de las semillas. Pues las semillas redondas no presionan sin ceder su forma, contra las que tienen enfrente, sino que ceden y son empujadas en los espacios entre tres o cuatro de las semillas que tienen en derredor. En colmenas, sin embargo, la razón es diferente. Pues las abejas no están apelotonadas de manera confusa como las semillas de una granada, sino que se alinean intencionadamente en posición de batalla, todas sus cabezas señalando en un sentido o el contrario, mientras empujan a la vez contra los fondos de las celdillas. Si la forma de la colmena fuese el resultado de un apelotonamiento, las celdillas tendrían que cubrir a las abejas a partir del endurecimiento de sus secreciones viscosas, como las conchas crecen alrededor de los caracoles en espiral. Pero está claro que las propias abejas construyen sus panales, levantando toda la estructura desde los cimientos." (Johannes Kepler, *El copo de nieve de seis picos. Un regalo de Año Nuevo*, [6].)

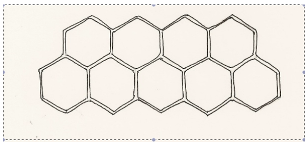
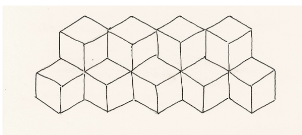
Se cree que fue la lectura de *Colección*, escrita por Pappus de Alejandría ([7], [8], [9]), lo que despertó la curiosidad de Kepler por la estructura del panal de miel. Pappus comienza el Libro V con el siguiente prefacio.

"La divinidad, querido Megetius, ha otorgado a los hombres la major y más perfecta noción de sabiduría en general y de ciencia matemática en particular, y sólo otorgó parcialmente este privilegio a los animales. A los hombres, al haberles dotado de razón, les concedió que hiciesen todo a la luz de la inteligencia y la demostración, y a los otros seres vivos, aunque les negó la razón, les concedió el que consiguiesen todo lo que les fuese vitalmente necesario gracias a un cierto instinto natural. La existencia de este instinto puede ser observado en muchas otras criaturas vivas, pero sobre todo en abejas. Son realmente admirables, en primer lugar, su disciplina y sumisión a las reinas que gobiernan en su estado. Pero mucho más admirable todavía es su emulación, la limpieza con que llevan a cabo la recolección de la miel, y el planificado y maternal cuidado con que la custodian. Probablemente porque saben que los dioses les han confiado la tarea de hacer llegar a parte selecta de la humanidad su ambrosía, no les parece adecuado dejarla caer descuidadamente sobre el suelo, la madera, o cualquier otro material feo e irregular. En vez de ello, tras recolectar los dulces de las más hermosas flores que crecen sobre la tierra, construyen para recoger la miel, los receptćulos que llamamos colmenas, (con cedillas) todas iguales, semejantes y contiguas unas a otras, y en forma hexagonal. Inferimos de la siguiente manera, que han llegado a ello en virtud de cierta planificación geométrica: necesariamente habrán pensado que las figuras habrían de ser contiguas unas a otras, es decir, teniendo paredes communes, de manera que ninguna materia extraña pudiese colarse en los intersticios entre ellas y profanar la pureza de su producto. Habiendo tres figuras con las que es posible recubrir exactamente el plano alrededor de un mismo punto, las abejas, debido a su astucia instintiva, eligen para la construcción del panal de miel la figura con el mayor número de ángulos, porque saben que contendrá más miel que cualquiera de las otras dos.

Las abejas, pues, conocen este hecho que les resulta muy útil, que el hexágono es mayor que el cuadrado y el triángulo[3] y dará cabida a más miel por el mismo coste de material utilizado en construir las distintas figuras. Nosotros, sin embargo, afirmando como afirmamos tener más inteligencia que las abejas, investigaremos un problema aún más profundo, el de que entre todas las figuras planas equiláteras y equiangulares de igual perímetro, la más extensa es la que tiene mayor número de ángulos, y la mayor figura plana entre todas las de igual perímetro es el círculo." (Pappus, *Colección Matemática*, Libro V, [7], pp. 237-239.)

Volvemos a la lectura de *El copo de nieve de seis picos.* Retomando el estudio de las colmenas donde lo dejase Pappus, Kepler analiza la geometría de las celdillas en profundidad.

"Si se les preguntase a los geómetras por el patrón seguido en la construcción de las celdillas de abejas, contestarán que un patrón hexagonal. La respuesta resulta evidente tras echar un simple vistazo a las aberturas o entradas, y a los lados que forman las celdillas [Fig. 4a]. Cada celdilla está rodeada por otras seis, y separada de la contigua por un pared compartida. Pero si observamos el fondo de cada celdilla, notaremos que desciende en un ángulo [diedro] obtuso formado por tres planos [Fig. 4b].

Fig. 4a   Fig. 4b

[3]Pappus se refiere aquí al hexágono regular, triángulo equilátero y cuadrado inscritos en una misma circunferencia, y a la noción de tamaño para figuras geométricas introducida por Euclides en *Elementos.* Los griegos utilizaban letras para expresar los números, lo que convertía el echar cuentas en una tarea bastante farragosa. Euclides evita introducir conceptos que requieran el uso de números –como área o volumen, por ejemplo–, y define una nueva manera de ser "iguales" dos figuras (planas o sólidas), que luego utilizará para comparar polígonos y poliedros: dos figuras son "iguales" si ambas pueden ser descompuestas en el mismo número (finito) de piezas más pequeñas congruentes dos a dos. Dicho de otra manera, ambas figuras son soluciones de un mismo rompecabezas de piezas planas o sólidas, por lo que podemos pasar de una figura a otra sin más que reordenar las piezas [5].

Este fondo (que podríamos llamar la quilla) se une a los seis lados de la celdilla por otros seis ángulos, tres superiores de tres lados intercalados con tres inferiores, de cuatro lados. Hay que añadir que la celdilla están distribuidas en dos capas, con las aberturas mirando direcciones opuestas, los fondos tocándose en un empaquetamiento cerrado; y la punta de cada quilla de una capa encajada en las puntas de tres de las quillas de la otra capa [Fig. 5]. Mediante este diseño, cada celdilla no sólo comparte seis paredes con las que le rodean en su misma capa, sino también tres planos al fondo con otras tres celdillas de la capa contraria.

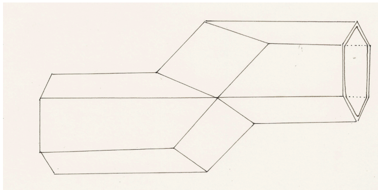


Fig. 5

Así cada abeja tiene nueve vecinas, con cada una de las cuales comparte una pared divisoria. Los tres planos de la quilla son idénticos entre sí y su forma es lo que los geómetras llaman un rombo. Intrigado por estos rombos, empecé a investigar en geometría a ver si se podía construir, utilizando exclusivamente rombos, un sólido como los cinco sólidos regulares y los catorce sólidos de Arquímedes."

Los sólidos regulares son cuerpos convexos, con polígonos regulares iguales por caras y ángulos iguales en todos sus vértices. Hay una infinidad de polígonos regulares, pero sólo cinco sólidos regulares: el tetraedro, el cubo, el octaedro, el dodecaedro y el icosaedro. Se les conoce también como los sólidos platónicos, al ser las figuras que usó Platón en el modelo con que describe el universo en el diálogo *Timeo*. Los sólidos de Arquímedes son sólidos 'semi-regulares," formados por polígonos regulares de más de un tipo. Hay trece de ellos (no catorce, como afirma Kepler), y nos han llegado a través del Libro V de *Colección matemática* de Pappus.

"Los filósofos dicen que, con toda razón, el primero de los dioses dio al mundo la forma de una figura esférica, pues se trata de la más bella entre las que existen. Mencionan las propiedades naturales de la esfera y añaden, además, que es la mayor entre todas las figuras con su misma superficie. Todo aquello que afirman en relación con las esfera es obvio y no requiere ninguna demostración; pero cuando afirman que es más grande que otras figuras, los filósofos no lo demuestran y se limitan a afirmarlo; pero esto no es fácil aceptarlo sin un análisis más minucioso. Por ello, y como antes hemos demostrado que el círculo es la mayor entre las figuras poligonales con su mismo perímetro, intentaremos demostrar ahora que la esfera es la mayor entre las figuras sólidas regulares con su misma superficie; pero consideraremos sobre todo las que parecen regulares. No se trata tan solo de las cinco figuras que aparecen en el divino Platón, a saber: el tetraedro y el hexaedro, el octaedro y el dodecaedro y, en quinto lugar, el icosaedro, sino también por aquéllas, en número de trece, descubiertas por Arquímedes, que están limitadas por polígonos equilaterales y equiangulares pero no todos iguales[4].

[4]Los sólidos semi-regulare o arquimedianos quedan determinados por el número C_K de caras con

Está, de entrada, el octaedro, limitado por 4 triángulos y 4 hexágonos.

Además de este último, hay tres decatetraedros, el primero limitado por 8 triángulos y 6 cuadrados, el segundo por 6 cuadrados y 8 hexágonos, y el tercero por 8 triángulos y 6 octógonos.

Además de estos últimos, hay dos icohexaedros, el primero está limitado por 8 triángulos y 18 cuadrados, y el segundo por 12 cuadrados, 8 hexágonos y 6 octógonos.

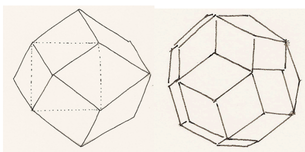
Además de estos últimos, hay tres triacontadoedros, el primero limitado por 20 triángulos y 12 pentágonos; el segundo por 12 pentágonos y 20 hexágonos, y el tercero por 20 triángulos y 12 decágonos.

Además de estos últimos, hay dos hexecontadoedros, el primero limitado por 20 triángulos, 30 cuadrados y 12 pentágonos, y el segundo por 30 cuadrados, 20 hexágonos y 12 decágonos.

Y además de todos ellos, el último es el enecontadoedro, limitado por 80 triángulos y 12 pentágonos." (Pappus de Alejandría, *Colección matemática*, Libro V, cap. XIX. [7], pp. 272-273.)

Retomamos la lectura de *El copo de nieve de seis picos* en el punto en que lo dejamos ([6] p. 43).

"Descubrí dos, uno relacionado con el cubo y el octaedro [Fig. 6a], el otro con el octaedro y el icosaedro [Fig. 6b] (el cubo mismo puede ser considerado un tercero, pues es como dos tetraedros unidos). El primero de estos sólidos es contenido por doce rombos, el otro por treinta. Sólo el primero comparte esta propiedad con el cubo: de la misma manera que ocho esquinas de ocho cubos unidas en un único punto llenarán el espacio sin dejar huecos, lo mismo ocurrirá con las esquinas del primer tipo de figura rómbica cuando cuatro de los ángulos obtusos (trilaterales) o seis de los cuadrilaterales se junten."

Fig. 6a  Fig. 6b

Es lógica la aparición de figuras con rombos por caras en semillas inicialmente esféricas que al crecer presionan unas contra otras, como en la granada o el guisante, observa Kepler. Pero si se tratase de esferas sólidas, ¿cómo habríamos de colocarlas para optimizar el espacio?, se pregunta (Fig. 7).

k lados, el número n_k de caras con k lados que pasan por cada vértice y los números C, A y V de caras, aristas y vértices del poliedro.

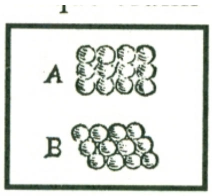


Fig. 7: Diagrama original de Kepler, [6] p. 54

> "Si las vas apilando unas sobre otras buscando el arreglo más compacto, habrás de ponerlas o bien en un arreglo cuadrado (Fig. A) o bien triangular (Fig. B)."

Esta afirmación, para la que Kepler no ofrece demostración, se conoce como La Conjetura de Kepler o *del frutero*. Con el tiempo llegó a ser tan importante como para formar parte de uno de los veintitrés problemas que, según la lista elaborada por Hilbert en 1900, marcarían el camino de la investigación matemática a lo largo del siglo 20. Los intentos por demostrar la Conjetura de Kepler dieron lugar a mucha, y muy buena, matemática a través de los siglos, hasta que en 1998 Thomas Hales dió con una demostración que, en lo que a demostraciones matemáticas se refiere, supone un caso excepcional. El argumento de Hales tiene una parte teórica publicada en 2005 ([4]), y otra computacional, tan técnica, que tras años estudiándola los expertos no supieron qué hacer con ella y acordaron incluirla como enlace en la página web de la revista[5]. El famoso diagrama con que Kepler ilustró su conjetura (Fig. 7), trae a la memoria otro diagrama igualmente famoso: el que utilizó Albert Einstein para ilustrar su modelo del universo (Fig. 8).

En el contexto de su trabajo en relatividad, Einstein propuso en 1917 un modelo de universo basado en la 3-esfera ([2]) que cuatro años más tarde describió para un público general en la conferencia *Geometría y experiencia*[6]. En la primera parte de su intervención, Einstein explica qué es lo que, en la teoría de la relatividad, sugiere considerar como modelo de universo un continuo espacio-tiempo cuatridimensional con métrica riemanniana. A continuación, da instrucciones detalladas de cómo llegar a una representación mental de tal continuo (de cuatro dimensiones, finito e ilimitado).

> "¿Podemos visualizar un universo finito pero ilimitado? La respuesta habitual a esta pregunta es 'no", pero esta no es la respuesta correcta. El propósito de las siguientes reflexiones es demostrar que la respuesta tendría que ser "sí" Quiero demostrar que se puede ilustrar sin gran dificultad la teoría de un universo finito, con una representación mental a la que, con un poco de práctica, pronto nos acostumbraremos.
>
> ¿Qué queremos expresar cuando decimos que nuestro espacio es infinito? Tan sólo que podríamos colocar cualquier cantidad de cuerpos de igual tamaño, unos

[5] Detalles sobre la historia de esta demostración pueden encontrarse en https://ztfnews.wordpress.com/2014/08/14/completada-la-prueba-formal-de-la-conjetura-de-kepler/

[6] Pronunciada ante la Academia Prusiana de Ciencias el 27 de enero de 1921 y recogida en su libro *Ideas y opiniones*, [3]

junto a otros, sin llegar a cubrir el espacio.[...] Veamos ahora un ejemplo de un continuo bidimensional que es finito pero sin límites. Imaginemos la superficie de un gran globo y una cantidad de pequeños discos de papel, todos del mismo tamaño. Colocamos uno de los discos en cualquier lugar sonbe la superficie del globo. Si desplazamos el disco a voluntad sobre la superficie del globo, en ningún momento del recorrido tropezaremos con un límite. Decimos, por lo tanto, que la superficie esférica del globo es un continuo sin límites.

Es más, la superficie esférica es un continuo finito. Porque si pegamos los discos de papel sobre el globo de manera que ningún disco se superponga a otro, la superficie de la esfera llegará a estar tan cubierta que será imposible colocar otro disco. Esto significa exactamente que la superficie esférica del globo es finita en relación a los discos de papel.

Y, más aún, la superficie esférica es un continuo no euclídeo de dos dimensiones, es decir, que las leyes de posición de figuras rígidas en ella no concuerdan con las del plano eucídeo. Esto puede demostrarse de la siguiente forma. Tomamos un disco y lo rodeamos, en círculo, por otros seis discos, cada uno de los cuales debe estar rodeado, a su vez, por otros seis discos y así sucesivamente (Fig. 8).

Fig. 8: Diagrama original, [3], p. 237

Si llevamos a cabo esta construcción sobre una superficie plana, obtendremos un patrón ininterrumpido con cada disco tocando otros seis, con excepción de los que estén en la parte externa. En un primer momento, esta construcción también promete éxito sobre una superficie esférica, y cuanto menor sea el radio del círculo en proporción al de la esfera, tanto más prometedora parecerá. Pero según se avanza en la construcción va resultando más patente que la colocación interrumpida de los discos en la manera indicada no es posible, como sí lo era en el caso de la geometría euclídea del plano. Es de esta manera que las criaturas que no puedan abandonar la superficie esférica, y no puedan siquiera echar una ojeada fuera de la superficie esférica hacia el espacio tridimensional, podrían llegar a descubrir, por simple experimentación con los discos, que su “espacio” bidimensional no es eucídeo, sino esférico.

Según los últimos resultados de la teoría de la relatividad, es probable que nuestro espacio tridimensional sea también aproximadamente esférico, esto es, que las leyes de localización en él de cuerpos rígidos no estén dadas por la geometría euclídea, sino en forma aproximada por la geometría esférica, si consideramos partes del espacio que sean suficientemente extensas. Este es el momento en que la imaginación del lector vacila. “Nadie puede imaginar semejante cosa”, exclama indignado. “Una cosa así puede ser dicha, pero no puede ser pensada. Soy capaz de imaginar bastante bien una superficie esférica, pero nada análogo a ella en tres dimensiones. Debemos intentar superar esta barrera mental y el lector paciente verá que no se trata de una tarea especialmente difícil. A tal fin, empezaremos por centrar nuestra atención, una vez más, en la geometría de las superficies esféricas bidimensionales. En la figura adjunta (Fig. 9), sea K la superficie esférica, que toca en S al plano E, representado, por facilitar la presentación, como una superficie limitada. Sea L un disco sobre la superficie esférica.

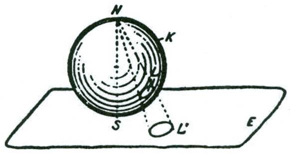


Fig. 9: Diagrama original, [3], p. 238

Ahora imaginemos que en el punto N de la superficie esférica, diametralmente opuesto a S, existe un punto luminoso que proyecta sobre el plano E una sombra L' del disco L. Todo punto sobre la esfera tiene su sombra en el plano. Si movemos el disco sobre la esfera K, también se moverá su sombra L' sobre el plano E. Cuando el disco L está centrado en S, coincide con su sombra. Si se mueve hacia arriba sobre la superficie esférica, alejándose de S, la sombra L' sobre el plano también se moverá hacia la parte externa del plano, alejándose de S y haciéndose más y más grande a medida que se aleje del punto citado. Según el centro del disco L se va aproximando al punto luminoso N, la sombra se va desplazando hacia el infinito y se hace infinitamente grande.

Ahora planteamos la siguiente pregunta: ¿cuáles son las leyes de posición de los discos-sombra L' sobre el plano E? Es evidente que son las mismas que las leyes de posicición de los discos L sobre la superficie esférica, dado que cada figura original sobre K tiene su figura-sombra correspondiente sobre E. Si dos discos se tocan sobre K, sus sombras también se tocarán sobre E. La geometría de las sombras sobre el plano, concuerda con la geometría de los discos sobre la esfera. Si denominamos a los discos-sombra figuras rígidas, entonces la geometría esférica es válida también sobre el plano E para estas figuras rígidas. En particular, el plano es finito con respecto a los discos-sombra, puesto que sólo un número finito de sombras tienen cabida en el plano.

En este punto alguien dirá: "Eso no tiene sentido. Los círculos proyectados no son figuras rígidas. Sólo tenemos que ir moviendo una regla de treinta centímetros por el plano E, para convencernos de que, a medida que se van alejando del punto S hacia el infinito, las sombras del plano van aumentando de tamaño constantemente".

Pero, ¿qué sucedería si la regla de treinta centímetros se comportase sobre el plano E de la misma manera que los discos-sombra L'? En tal caso sería imposible demostrar que las sombras crecen en tamaño a medida que se alejan de S y tal afirmación no significaría nada. En realidad, lo único que se puede afirmar objetivamente sobre los discos-sombra es lo siguiente: estos discos están relacionados entre sí en el sentido de la geometría euclídea, exactamente de la misma manera en que lo están los discos rígidos sobre la superficie esférica. Debemos tener muy claro que nuestra afirmación con respecto al crecimiento de los discos-sombra a medida que se van alejando de S hacia el infinito, no tiene significación objetiva en sí, en cuanto seamos incapaces de comparar los discos-sombra con los cuerpos rígidos euclídeos que se mueven sobre el plano E. Con respecto a las leyes de posición relativa de las sombras L', el punto S no tiene más privilegios especiales sobre el plano que sobre la superficie esférica.

La representación de la geometría esférica sobre el plano que acabamos de ver es importante para nosotros, porque puede ser transferida fácilmente al ámbito tridimensional. Imaginemos un punto S en nuestro espacio y una gran cantidad de pequeñas esferas L' que podemos hacer coincidir unas con otras. Sin embargo, estas esferas no han de ser rígidas en el sentido de la geometría euclídea; su radio ha de aumentar (en el sentido de la geometría euclídea) cuando se muevan hacia el infinito alejándose de S, y ese aumento ha de producirse según la misma ley que determinaba el crecimiento de los radios de los discos-sombra L' sobre el plano.

Una vez tenemos una imagen mental vívida del comportamiento geométrico de nuestras esferas L', supongamos que en nuestro espacio no existen cuerpos rígidos en el sentido de la geometría euclídea, sino sólo cuerpos que se comportan como las esferas L', y contaremos con una imagen clara del espacio esférico tridimensional o, mejor aún, de la geometría esférica tridimensional. Ahora son las esferas, las que deben ser llamadas "rígidas". Como en el caso de los discos-sombra en el plano E, el aumento de su tamaño a medida que se van alejando de S no puede ser detectado con reglas de medición, porque los patrones de medida se comportan del mismo modo que las esferas. Nuestro espacio es finito porque, debido a su "crecimiento", sólo un número finito de esferas puede tener cabida en el espacio.

De esta manera, apoyándonos en la práctica en pensar y visualizar que nos da la geometría euclídea, hemos construido una representación mental de la geometría esférica. Sería posible, sin gran dificultad, llevar a cabo construcciones imaginarias especiales que diesen mayor profundidad y fuerza a estas ideas. Pero mi único objetivo hoy ha sido demostrar que la facultad humana de visualización no está obligada en absoluto a capitular ante la g)eometría no euclídea." (Albert Einstein, [3], pp. 235-240.)

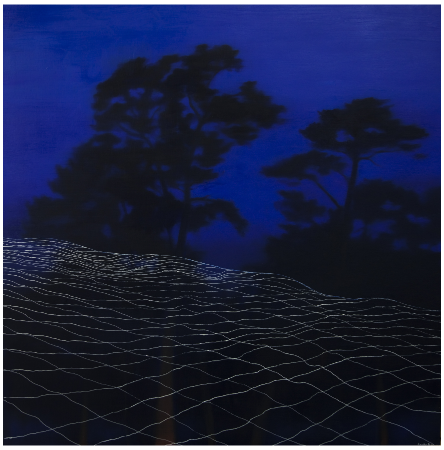


Fig. 10: Amadeo Olmos 2010

En la introducción a su edición de algunos textos de Riemann ([11], p. XCIII), José Ferreirós explica de donde surgió la idea de Einstein de considerar un continuo cuatridimensional finito pero ilimitado:

Es bien sabido –escribe Ferreirós–, *que la geometría de Riemann se inscribe en uno de los capítulos más famosos de la historia de la matemática, el desarrollo de las geometrías no euclídeas, y que constituyó un preparativo esencial para la teoría relativista de la gravitación. Puede decirse, abusando de una expresión famosa, que aquí nos encontramos con tres gigantes subidos los unos a hombros de otros: Riemann sobre las espaldas de Gauss, y Einstein encima de aquél. A propósito de este desarrollo de ideas, escribía precisamente Einstein en 1922*:

> "Los conocimientos matemáticos que posibilitaron el establecimiento de la teoría de la relatividad general, hemos de agradecérselos a las investigaciones geométricas de Gauss y Riemann. El primero investigó en su teoría de superficies las propiedades métricas de una superficie inmersa en un espacio euclídeo tridimensional, y demostró que éstas pueden ser descritas mediante conceptos que sólo guardan relación con la propia superficie, y no con la inmersión [en el espacio] . . . Riemann extendió la línea de pensamiento gaussiana a continuos de cualquier número de dimensiones, y previó el significado físico de esta generalización de la geometría de Euclides con visión profética. . . . las leyes naturales sólo adoptan su forma lógica más satisfactoria cuando son expresadas como leyes del continuo cuatridimensional espacio-temporal."

Veamos la descripción que Riemann dio de las variedades y, en particular, las variedades de cuatro dimensiones, a que hace referencia Einstein.

> "Atribuimos a un objeto variación continua cuando es posible una transición continua de una determinación del mismo a otra. La totalidad de las determinaciones (o también parte de ellas, entre las que sea posible una transición continua) constituyen entonces una variedad extensa continua, y cada una de ellas se llama punto de esa variedad.
>
> [. . .] El concepto de variedad de múltiples dimensiones subsiste independientemente de nuestras intuiciones espaciales. El espacio, el plano, la línea, son sólo los ejemplos más intuitivos de una variedad de tres, dos o una dimensión. Aún sin tener la menor intuición espacial, podríamos desarrollar toda la geometría. Aunque podríamos deducir analíticamente todas las proposiciones sobre variedades de más de tres dimensiones, sería con mucho preferible el basar la teoría de las variedades de más dimensiones directamente sobre la geometría. Quiero desarrollar esto solamente para las variedades de 4 dimensiones.
>
> [. . .] Una variedad de 4 dimensiones es aquí, pues, algo que contiene en sí infinitos espacios. Pero nunca podemos contemplar más que lo que está en un mismo espacio. Cuando diga, en lo sucesivo, que dos construcciones no están en el mismo espacio, quiere ello decir que no puedo contemplar ambas a la vez, no puedo hacerme absolutamente ninguna imagen de su lugar recíproco, sólo puedo extraer conclusiones lógicas a partir de las premisas que hago sobre ellas, y es un requisito esencial para la comprensión de lo que sigue, el atenerse sólo a estas conclusiones lógicas.

> [. . .] Queremos ocuparnos de una espacialidad de cuatro dimensiones. Se trata de una espacialidad continua de mayor extensión que todo el espacio infinito. Ciertamente, con nuestra intuición no podemos abarcar más que todo el espacio infinito, pero podemos desplazarnos a través de diversos espacios, uno tras otro. Quiero decir con ello: me imagino primero todo el espacio infinito, después me imagino otra vez todo el espacio infinito, pero pensando que todos los puntos que ahora contemplo son distintos de los puntos que contemplaba antes." (Bernhard Riemann, *Fragmentos sobre variedades y geometría*, en [11], p. 93-95.)

La imagen mental de la espacialidad de cuatro dimensiones que describe Riemann en el último párrafo, es una analogía de la representación cartográfica del globo terráqueo como dos hemisferios planos circulares. Cada uno de los círculos representa la mitad de la esfera terrestre, y sus circunferencias, el círculo máximo terrestre común a ambos. Si desde cualquier punto del interior de cualquiera de los dos hemisferios nos acercamos a su circunferencia, nos encontraremos automáticamente en el punto correspondiente del segundo hemisferio, y podremos abarcar el interior de ambos con la vista. Imaginemos de nuevo que tenemos delante dos círculos, pero esta vez, sin embargo, cada uno de ellos representa una esfera. La Tierra está colocada en el centro de una de ellas, que tiene en su interior todo lo que nuestra civilización alcanza a ver. A continuación, imaginamos que una civilización lejana, más allá del alcance de nuestra visión, está situada en el centro de la segunda esfera, cuyo interior contiene todo lo que tal civilización alcanza a ver. Hay varias posibilidades teóricas: que las esferas se encuentren muy lejos una de otra con mucho universo entre ellas, que tengan partes en común con galaxias observables para ambas civilizaciones o, según propone Riemann, que compartan su superficie y juntas constituyan todo el universo, como los dos hemisferios del mapa de la Tierra comparten su circunferencia y juntos forman el globo terráqueo. Si desde cualquier punto del interior de cualquiera de las dos esferas nos acercamos a la superficie común, nos encontraremos automáticamente en el correspondiente punto de la segunda esfera, y podremos abarcar el interior de ambas con la vista. En 1997, el físico Mark Peterson llamó l atención de la comunidad científica sobre el hecho de que la primera descripción escrita conocida del universo como dos esferas que comparten su superficie, y de la experiencia de poder abarcar simultáneamente, desde cualquier punto del borde común a ambas, el interior de las dos, la dio el poeta Dante Alighieri en el siglo XIV (*La Divina Comedia. Paraíso*, [1], p. 555 y ss.).

> 'Tan excelsa y tan viva en esa esfera, / y tan igual, es cada parte, que no entiendo / en cuál Beatriz el sitio me escogiera.
> [...] La natura del mundo, que está quieta / en su centro, mas todo en torno mueve, / comienza aquí desde su propia meta. (Canto XXVII, 100-102, 106-108)
> Como del cirio que le está alumbrando / detrás, en el espejo advierte el fuego / quien no lo ve ni en él iba pensando,
> vi un punto que irradiaba una clareza / tan aguda, que al ojo que la enfoca / le obliga a que se encierre su agudeza (Canto XXVIII, 4-6, 16-18)"

En la última parte de su ensayo, Kepler vuelve a los copos de nieve. Todos los poliedros considerados por los griegos habían sido, como la mayoría de los copos de nieve, convexos.

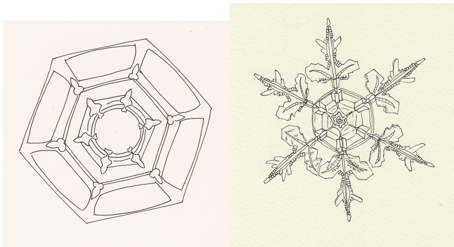


Fig. 11

> "Mientras escribo esto, ha empezado a nevar, y mucho más copiosamente que hace un rato. He estado examinando con detenimiento los pequeños copos. Bien, han estado cayendo todos ellos con un patrón radial, pero de dos tipos. Algunos muy pequeños y con una cantidad indefinida de púas insertadas por todas partes, son de formas sencillas sin plumas ni estrías y muy finas, y tienen un glóbulo un poco más grande en el centro. Estos forman la mayoría. Pero salpicados entre ellos aparecen los copos del segundo tipo, las estrellas con seis plumas."

La existencia de copos de nieve estrellados, llevó a Kepler a investigar la posibilidad de construir poliedros que, como estos copos, fuesen cóncavos. Acabó re-descubriendo dos, el pequeño dodecaedro estrellado (que podemos ver en los mosaicos de Ucello de la Catedral de San Marco, 1425-1430) y el gran dodecaedro estrellado (descrito ya en 1568 por Wenzel Jamnitzer). El descubrimiento de Kepler dió lugar a una nueva pregunta: ¿cuántos poliedros regulares, cóncavos o convexos, hay en total? En 1809 Poinsot encontró dos más, el gran dodecaedro y el gran icosaedro y Cauchy demostró en 1813 que sólo hay nueve: los cinco sólidos platónicos, convexos, y los cuatro, cóncavos, de Kepler y Poinsot.

Como broche de oro, Kepler nos regala, escondida entre sus reflexiones sobre la diferencia entre los patrones hexagonales y pentagonales que encontramos en la naturaleza, la primera descripción escrita, que se conoce, de la relación entre la serie numérica introducida por Fibonacci en *Liber Abaci* (1202) y la proporción divina ([6], p. 67). Denominada por los griegos como media y extrema razón (Euclides, Elementos, Libro VI) y por los lectores de Matila Ghyka (1931) como el número de oro o proporción aúrea, su valor numérico es $\frac{1+\sqrt{5}}{2}$.

> "La estructura de cada uno de estos sólidos, como la del pentágono, no puede ser construida sin la proporción que los geómetras modernos denominan "divina." Está ordenada de tal manera que los dos términos menores en una proporción continua producen el tercero; y, sucesivamente, cualesquiera dos términos adyacentes suman el término siguiente, hasta el infinito, pues la misma proporción se mantiene siempre.
>
> Es imposible dar un ejemplo numérico perfecto. Cuanto más nos alejemos de la unidad, más perfecto será nuestro ejemplo. Sean los términos menores 1 y 1,

que debes pensar como distintos. Súmalos, y hacen 2. Añade el mayor 1 para hacer 3. Suma 2 y hacen 5. Suma 3 y hacen 8. Suma 5 y hacen 13. Suma 8 y hacen 21. Como 5 es a 8, así es, aproximadamente, 8 a 13; como 8 es a 13 así es, aproximadamente, 13 a 21. Y así sucesivamente, hasta la eternidad."

Hay quienes, quizás atraídos por su divina proporción, consideran la serie de Fibonacci como una escalera de oro que nos acerca a la eternidad, y la geometría de los pentágonos y hexágonos el no va más y sustento estructural del universo. Los matemáticos, sin embargo, hacen suyas la palabras finales de Kepler,

"Pero la facultad formativa de la tierra no se restringe a una única figura, sino que practica con maestría toda la geometría."

y equipados con su caja de herramientas siguen explorando el universo en busca de estructuras nuevas por las que trepar. Como José María Montesinos Amilibia.

References

[1] Dante Alighieri: *Divina Comedia.* (1307-1321), traducción de Ángel Crespo, Editorial Planeta 1990.

[2] A. Einstein: *Cosmological considerations on the general relativity theory.* En *The Principle of Relativity, a collection of original memoirs on the special and general theory of relativity,* H. A. Lorentz, A. Einstein, H. Minkowski, and H. Weyl, with notes by A. Sommerfeld, 177–188, translated by W. Perrett and G. B. Jeffery. Published by Methuen and Co. Ltd., London, 1923; reprinted in 1990; also published unaltered and unabridged by Dover Publications, Inc., in 1952.

[3] A. Einstein: *Ideas and opinions* (1954). Crown Publishers 1960.

[4] T. Hales: *A proof of the Kepler conjecture.* Annals of Mathematics 162-3 (2005), 1065-1185.

[5] R. Hartshorne: *Euclid and Beyond.* Springer-Verlag, 2000.

[6] Johannes Kepler: *The Six Cornered Snow Flake. A New Year's Gift* (1611). Edición bilingűe latín/inglés, Paul Dry Books 1910, ilustraciones y notas matemáticas de Capi Corrales Rodrigáñez.

[7] Pappus de Alejandría: *La Collection Mathématique (340) Livre V.* Tome Premier, 237–364, A. Blanchard 1982.

[8] Pappus de Alejandría: *Pappus on the sagacity of bees.* En *The History of Mathematics. A Reader,* J. Fauvel y J. Gray (eds.), MacMillan 1987, p. 211.

[9] Pappus de Alejandría: *The Collection. Book V. Preface on the Sagacity of Bees.* En *A History of Greek Mathematics* (1921) Vol. II, S.T. Heath, Dover Publications 1981, p. 389.

[10] M. A. Peterson: *Dante and the 3-sphere.* American Journal of Physics 47, 12 (1979) 1031–1035.

[11] J. Ferreirós: *Riemanniana Selecta.* Ediciones del CSIC, Madrid 2000.

A note on regular branched foldings

Antonio F. Costa [*]

Departamento de Matemáticas Fundamentales
Facultad de Ciencias
UNED
28040 Madrid, Spain
acosta@mat.uned.es

Dedicated to my friend and professor José María Montesinos.

ABSTRACT

In this note we introduce branched foldings between compact surfaces using a local description. We remark that if some branched folding has some particular local structure then it will be non-regular.

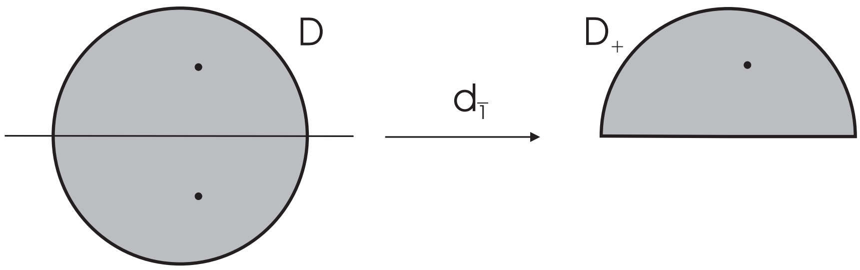
2010 Mathematics Subject Classification: 14E20, 57M12, 57N05.
Key words: foldings, branched coverings, orbifolds, Klein surfaces.

1. Introduction

For more than thirty years I had the privilege of enjoying the friendship of professor José María Montesinos; along this time we had the opportunity of talking about many mathematical and no-mathematical problems. His personal points of view provide any subject one can think of with new light and original ideas. This is illustrated by this very work; indeed it has its origins on the joint preparation of some lectures for a Master of Mathematics of the Complutense University. We then discovered a new fact on coverings of 2-orbifolds which I, after spending several years working on that field, had never thought of.

Branched foldings were introduced in [4]. In this article we consider branched foldings between compact surfaces. These maps present the topological type of either coverings between compact 2-orbifolds (see [5] and [3]) or morphisms between compact Klein surfaces (see [1]). We shall introduce branched foldings using a local description.

[*]Author partially supported by project MTM2014-55812



Orbifolds appear as an useful tool to study properly discontinuous actions of groups on manifolds (see [2]), and the more important examples of orbifold coverings are the projections from manifolds to orbit spaces of group actions. In the theory of Klein surfaces the regular morphisms are specially important as well.

In this article we shall claim a global property (in a negative way) from a local behavior. Concretely: if some branched folding has some particular local structure then it will be non-regular, i. e. it is not the projection from a surface to a quotient by a group action.

2. The local models

All the surfaces that we shall consider are orientable or non-orientable, with or without boundary and always compact.

We start with some notation. We note:

$$D = \{z \in \mathbb{C} : |z| \leq 1\}$$

and

$$D_+ = \{z = a + bi \in \mathbb{C} : |z| \leq 1, \quad b \geq 0\} :$$

Given a map $f : S_1 \to S_2$ between surfaces, an automorphism of f is a homeomorphism $\phi : S_1 \to S_1$ such that $f \circ \phi = f$. The automorphisms of f form a group with composition. We denote the group of automorphisms of f by $\mathrm{Aut}(f)$.

The following maps are the "local building blocks" for the definition of branched foldings.

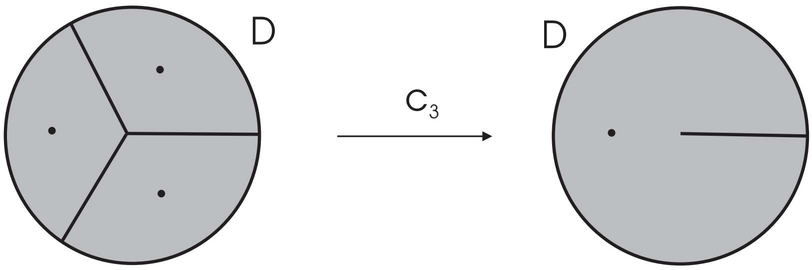
There are four types of LOCAL BRANCHED FOLDINGS:

I) The LOCAL DIHEDRAL MAP $d_{\bar{1}} : (D, 0) \to (D_+, 0)$, $z = a + bi \longmapsto z' = a + |b|\, i$.

Remark 2.1 *1.* $\mathrm{Aut}(d_{\overline{1}})$ *is the dihedral group* $\mathfrak{D}_{\overline{1}}$ *of order* 2*, generated by the homeomorphism* $z \longmapsto \overline{z}$ *of* D*.*

2. $\mathrm{Aut}(d_{\overline{1}})$ *acts transitively on* $d_{\overline{1}}^{-1}(z)$ *for every* $z \in D_+$*.*

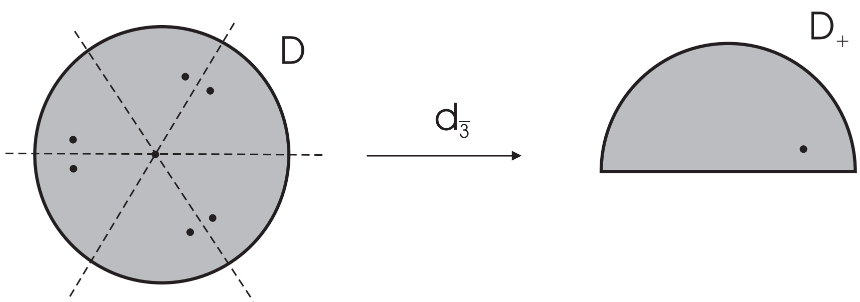
II) The LOCAL N-CYCLIC MAP $c_n : (D, 0) \to (D, 0)$, $z \longmapsto z^n$, $n > 1$.



Remark 2.2 *1.* $\mathrm{Aut}(c_n)$ *the cyclic group* $\mathfrak{C}_n$*, and so* c_n *is the quotient map under the rotational action of* $\mathfrak{C}_n$ *on* D *generated by* $z \longmapsto e^{2\pi i/n}z$*.*

2. $\mathrm{Aut}(c_n)$ *acts transitively on* $c_n^{-1}(z)$ *for all* $z \in D$*.*

III) The LOCAL $\overline{n}$-DIHEDRAL MAP $d_{\overline{n}} : (D, 0) \to (D_+, 0)$ is the composition of the local n-cyclic map with the local dihedral map, $d_{\overline{n}} = d_{\overline{1}} \circ c_n$.



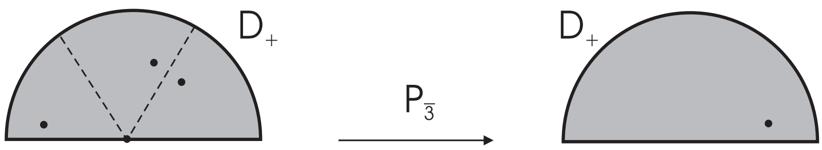
Remark 2.3 *1.* $\mathrm{Aut}(d_{\overline{n}})$ *is the dihedral group* $\mathfrak{D}_{\overline{n}}$ *of order* $2n$ *on* D*, generated by the homeomorphisms* $z \longmapsto e^{2\pi i/n}z$ *and* $z \longmapsto \overline{z}$*.*

2. $\mathrm{Aut}(d_{\overline{n}})$ *acts transitively on* $d_{\overline{n}}^{-1}(z)$ *for all* $z \in D_+$*.*

IV) the LOCAL $\overline{n}$-FOLDING MAP $p_{\overline{n}} : (D_+, 0) \longmapsto (D_+, 0)$ is the diagonal map that makes commutative the following diagram

$$\begin{tikzcd} D \arrow[r, "d_{\overline{n}}"] \arrow[d, "d_{\overline{1}}"'] & D_+ \\ D_+ \arrow[ur, "p_{\overline{n}}"'] & \end{tikzcd}$$

where the horizontal map is the local $\overline{n}$-dihedral map and the vertical map is the local $\overline{1}$-dihedral map.



Remark 2.4 *1.* $\mathrm{Aut}(p_{\overline{n}}) = \{1\}$ *if n is odd and $Aut(p_{\overline{n}}) = \mathfrak{D}_{\overline{1}}$ if n is even.*
2. (up if $n = 2$) $\mathrm{Aut}(p_{\overline{n}})$ *does NOT act transitively on $p_{\overline{n}}^{-1}(z)$ for every $z \in D_+$.*

3. Branched foldings between surfaces

Let us establish some terminology:

Definition 3.1 *Two maps between surfaces $f_1 : U_1 \to V_1$ and $f_2 : U_2 \to V_2$ are* TOPOLOGICALLY EQUIVALENT *if there are homeomorphisms $h : U_1 \to U_2$ and $g : V_1 \to V_2$ such that $g \circ f_1 = f_2 \circ h$.*

The main definition of this work is a particular case of the definition 3.5 of [4]:

Definition 3.2 *Let S_1 and S_2 be two surfaces. A map $f : S_1 \to S_2$ is called a* BRANCHED FOLDING *if every $x \in S_2$ possesses an elementary neighborhood U_x. The neighborhood U_x is called* ELEMENTARY *iff*
1. Each connected component V_x of $f^{-1}(U_x)$ intersects $f^{-1}(x)$ in one point denoted y_{V_x}; and
2. The restriction

$$f|_V : (V_x, y_{V_x}) \to (U, x)$$

is either a homeomorphism or it is topological equivalent to one of the four local branched foldings.

Examples. 1. The natural map $F \to F/G$, where G is any finite group of simplicial automorphisms acting on a triangulated surface F, is a branched folding.

2. Let $f : O_1 \to O_2$ be a covering between compact 2-orbifolds and $|f| : |O_1| \to |O_2|$ be the topological map induced on the topological surfaces $|O_1|$ and $|O_2|$ underlying to the orbifolds O_1 and O_2. The map $|f|$ is a branched folding. A morphism between compact Klein surfaces being particular case of orbifold covering is a branched folding as well.

Definition 3.3 *If f is a branched folding and the group $\mathrm{Aut}(f)$ is transitive on all the fibers of f we say that f is* REGULAR.

Remark 3.1 *1. If $f : S_1 \to S_2$ is regular then f is the natural projection given by the action of the group $\mathrm{Aut}(f)$ on S_1.*

2. The maps $d_{\overline{1}}$, c_n, $d_{\overline{n}}$, $p_{\overline{2}}$ are examples of regular branched foldings and the maps $p_{\overline{n}}$, $n \neq 2$ are not regular

3. If $f : S_1 \to S_2$ is a regular branched folding and S_2' is a surface contained in S_2 then $f|_{f^{-1}(S_2')}$ is a regular branched folding.

This remark leads to the main claim of this work:

Proposition 3.1 *If $f : S_1 \to S_2$ is a branched folding such that locally it is topologically equivalent to $p_{\overline{n}}$, $n > 2$ then f is not regular.*

Remark 3.2 *The above proposition only has non-empty meaning if $f : S_1 \to S_2$ is a branched folding but not a branched covering. The hypothesis in the above proposition is local in the sense it refers to the behavior of f in a neighborhood of a point in S_1. For branched coverings we can not deduce the non-regularity from a local property.*

There are many non-local possible conditions to claim the non-regularity of branched foldings, for instance:

Proposition 3.2 *Let $f : S_1 \to S_2$ be a branched folding, $x \in S_2$ and U_x be an elementary neighborhood of x. If there are two connected components of $f^{-1}(U_x)$ topologically equivalent to two different types of local branched foldings then f is not regular.*

References

[1] N. L. Alling, N. Greenleaf: *Foundations of the theory of Klein surfaces.* Lecture Notes in Mathematics 219, Springer-Verlag, Berlin-New York, 1971.

[2] M. Boileau, S. Maillot, J. Porti: *Three-dimensional orbifolds and their geometric structures.* Panoramas et Synthèses [Panoramas and Syntheses], 15. Société Mathématique de France, Paris, 2003.

[3] J. M. Montesinos-Amilibia: *Classical tessellations and three-manifolds.* Universitext. Springer-Verlag, Berlin, 1987.

[4] J. M. Montesinos-Amilibia: *Branched folded coverings and 3-manifolds*. En *Mathematical contributions in honor of Juan Tarrés,* 295–315, UCM 2012.

[5] W. P. Thurston: *The geometry and topology of three manifolds*. Chapter 13. http:// library.msri.org/nonmsri/gt3m/PDF/13.pdf.

Partially flat surfaces solving k–Hessian perturbed equations

Gregorio Díaz* and **Jesús Ildefonso Díaz**†

Instituto de Matemática Interdisciplinar
Departamento de Matemática Aplicada
Universidad Complutense de Madrid
28040 Madrid, Spain
gdiaz@ucm.es jidiaz@ucm.es

To José María Montesinos in occasion of his 70th birthday: a deep teacher and a good colleague.

ABSTRACT

In this paper we study a free boundary arising when a kind of diffusion involving Hessian functions is placed in balance with an absorption term (zero order nonlinear term of the own solution u). The diffusion operator is the k^{th} elementary symmetric function of the eigenvalues of the Hessian matrix D^2u and the absorption is a real increasing function vanishing at the origin such that the $(k+1)$–root of its primitive is integrable near the origin.

The surface associated to this solution has a strictly convex part and some flat sides. The junction between both regions of the surface behaves like a free boundary due to the degeneracy of the elliptic leading part of the equation on this interface.

2010 Mathematics Subject Classification: 35J60, 35J96, 35R35, 53C45.
Key words: Hessian equations, surfaces with flat sides, free boundary, degenerate elliptic operators..

*Supported by the UCM Research Group MOMAT (Ref. 910480) and by the projects MTM2011-26119 and MTM2014-57113, (DGISPI, Spain).

†Supported by the UCM Research Group MOMAT (Ref. 910480) and by the projects MTM2011-26119 and MTM2014-57113, (DGISPI, Spain).

1. Introduction

Given $u \in \mathcal{C}^2(\Omega)$ we denote by $\lambda(\mathrm{D}^2u) = \big(\lambda_1, \ldots, \lambda_{\mathrm{N}}\big) \in \mathbb{R}^{\mathrm{N}}$ the eigenvalues of D^2u. Then we consider the k^{th} elementary symmetric function

$$\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2u\big)\big] = \sum_{1\le i_1<i_2<\cdots<i_k\le \mathrm{N}} \lambda_{i_1}\cdots\lambda_{i_k}, \tag{1.1}$$

where Ω is an open bounded set of $\mathbb{R}^{\mathrm{N}}$, $\mathrm{N} > 1$. Obviously, k is an integer number taking value in $[1,\mathrm{N}]$. Therefore, the case $k = 1$ corresponds to the Laplacian operator $\mathcal{S}_1\big[\lambda\big(\mathrm{D}^2u\big)\big] = \Delta u$ while it is a fully nonlinear operator in the other choices of k. For example, the choice $k = 2$ leads to $\mathcal{S}_2\big[\lambda\big(\mathrm{D}^2u\big)\big] = \dfrac{1}{2}\Big(\big(\Delta u\big)^2 - |\mathrm{D}^2u|^2\Big)$ and $k = \mathrm{N}$ leads to the Monge–Ampere operator $\mathcal{S}_{\mathrm{N}}\big[\lambda\big(\mathrm{D}^2u\big)\big] = \det \mathrm{D}^2u$. Such kind of eigenvalues products are of relevance in the study of calibrate geometry (see [15]).

We note that a kind of Strong Maximum Principle holds for the *admissible* solutions of

$$\mathcal{S}_k[\lambda(\mathrm{D}^2u)] \ge 0 \quad \text{in } \Omega.$$

Therefore they can not assume an interior maximum or minimum value unless are constant solutions. The main goal of this paper is to prove that this kind of positivity information can be violated generating a dead core in Ω whenever the Hessian function is balanced against suitable absorptions. This paper will extend our previous work [8] dealing with the perturbed Monge-Ampere equations to the case of arbitrary k–Hessian equations for any available k.

More precisely, we focus our the attention on the equation

$$\mathcal{F}(\mathrm{D}^2u, \mathrm{D}u, u) = 0 \quad \text{in } \Omega, \tag{1.2}$$

for the fully nonlinear operator

$$\mathcal{F}(\mathrm{D}^2u, \mathrm{D}u, u) \doteq -\mathcal{S}_k[\lambda(\mathrm{D}^2u)] + \eta g\big(|\mathrm{D}u|\big) f\big(u - h\big), \quad u \in \mathcal{C}^2 \tag{1.3}$$

with $\eta > 0$, $g \in \mathcal{C}^1(\mathbb{R}_+; \mathbb{R}_+)$, $f \in \mathcal{C}(\mathbb{R})$ and $h \in \mathcal{C}(\Omega)$. The operator $\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2u\big)\big]$ is only elliptic when $u \in \mathcal{C}^2(\Omega)$ is k–admissible [3], namely the eigenvalues $\lambda(\mathrm{D}^2u)$ lie in the open symmetric convex cone, $\overline{\Gamma}_k$, in $\mathbb{R}^{\mathrm{N}}$ with vertex at the origin (see Section 2), then some compatibility is required on the structure of the equation (1.2) when is restricted to k-admissible solutions. In fact, the operator is degenerate elliptic on the symmetric definite non–negative matrices (see the comments in Section 2). As it will be proved in Theorem 3.2 (see also Remark 3.3), the compatibility is based on

$$h \text{ is locally } k\text{-admissible on } \overline{\Omega} \text{ and } h \le u \text{ on } \partial\Omega. \tag{1.4}$$

We emphasize that if f is too flat near the origin (see (1.7) below) and $u(x_0) > h(x_0)$ at some $x_0 \in \partial\Omega$ or $\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2h(x_0)\big)\big] > 0$ at some point $x_0 \in \Omega$ then $\mathcal{F}(\mathrm{D}^2u, \mathrm{D}u, u)$ is non-degenerate in path-connected open sets Ω (see Corollary 2).

The paper is organized as follows. In Section 2 we collect in a short way several comments on the notions of solutions. In Section 3 we obtain some weak maximum principles for the associated boundary value problem to (1.2) and get an existence of solutions result. Section 4 deals with the study of flat regions: we give some sufficient conditions for its occurrence as well as some estimates on its location. The consideration of unflat solutions is carried out in Section 5. The results can be considered, in some sense, as necessary conditions for the existence of flat solutions in terms of the zero order term of the equation.

One of the main consequence of the Weak Maximum Principle is the comparison result for which one deduces $h \leq u$ on $\overline{\Omega}$, provided (1.4) holds, i.e., h behaves as a kind of lower "obstacle" for the solution u (see Remark 3.3 below). Therefore, under (1.4) for any $\varphi \in \mathcal{C}(\partial\Omega)$, the boundary value proble considered in this papers is

$$\begin{cases} \mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u\big)\big] = \eta g\big(|\mathrm{D}u|\big) f\big(u-h\big) & \text{in } \Omega, \\ u = \varphi & \text{on } \partial\Omega, \end{cases} \tag{1.5}$$

where Ω is a bounded $(k-1)$–covex open set of $\mathbb{R}^{\mathrm{N}}$, $\mathrm{N} > 1$ (see below). We note that the usual restriction on the non negativity of the right hand side is here supplied by (1.4). We emphasize that since the right hand side of the equation needs not to be strictly positive in some region of Ω, the ellipticity of the Hessian function and the regularity $\mathcal{C}^2$ of solutions cannot be "a priori" guaranteed. The so-called "viscosity k–admissible solutions" or the "generalized k–admissible solutions" are adequate notions in order to weaken the non-degeneracy hypothesis on the operator. By using the Weak Maximum Principle and well known methods we prove, in Theorem 3.2, the existence of a unique generalized solution of (1.5). By a simple reasoning we obtain estimates on the gradient $\mathrm{D}u$. Bounds for the second derivatives $\mathrm{D}^2 u$ can be deduced from second order estimates (see Remark 3.3 below).

Since $h \leq u$ holds on $\overline{\Omega}$, the junction between the regions where $\{u = h\}$ and $\{h < u\}$ is a *free boundary*, thus it is not known a priori. This free boundary can be defined also as the boundary of the set of points $x \in \Omega$ for which $\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u(x)\big)\big] > 0$. Obviously, since the interior of the regions $\{u = h\}$ and $\{\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u\big)\big] = 0\}$ coincide, we must have $\mathcal{S}_k[\lambda(\mathrm{D}^2 h)] = 0$ in these interior region. The occurrence and localization of this free boundary is studied in Section 4 whenever $h(x)$ has *flat regions*

$$\mathrm{Flat}(h) = \bigcup_{\alpha}\{x \in \Omega : h(x) = \langle \mathbf{p}_\alpha, x\rangle + a_\alpha,\ \mathbf{p}_\alpha \in \mathbb{R}^{\mathrm{N}},\ a_\alpha \in \mathbb{R}\} \neq \emptyset,$$

where $\langle\cdot,\cdot\rangle$ denotes the Euclidean inner product in $\mathbb{R}^{\mathrm{N}}$. As it will be proved, the free boundary appears under two different types of conditions on the data: a precise behavior of the zeroth order term

$$\int_{0^+} \mathrm{F}(t)^{-\frac{1}{k+1}}\, dt < \infty \tag{1.6}$$

(or $0 < \mathrm{q} < k$ for $f(t) = t^{\mathrm{q}}$), where $\mathrm{F}(t) = \displaystyle\int_0^t f(s)ds$, and a suitable balance between the "size" of the regions of Ω where h is flat and the "size" of the data φ and h.

We shall show here how the theory on free boundaries (essentially the boundary of the support of $u - h$), developped for a class of quasilinear operators in divergence form, can be extended to the case of the solution of (1.2) inside of flat regions of h, where $u_h = u - h$ solves

$$\mathcal{S}_k[\lambda(\mathrm{D}^2 u_h] = \lambda g\big(|\mathrm{D}u|\big) f(u_h).$$

This kind of question has been extensively studied in the monography [9], mainly for the quasilinear p–Laplacian operator (see also [7] for fully nonlinear operators). In fact, the results of this paper were suggested in [9] and performed in [8] for the Monge–Ampere operator. The main existence criterion for the free boundary is strongly based on the condition (1.6). Clearly, it coincides with the corresponding main assumption used in [9] for the Laplacian operator. Since the strict k-admissibilty must be removed, a critical size of the data is required: the parameter η governs these kind of condition (see (4.30) below). For instance, the second mentioned balance for given φ and Ω is satisfied if η is large enough.

In Theorems 3 and 5 below we prove the occurrence of the free boundary and give some estimates on its localization. We also prove that if $h(x)$ growths moderately (in a suitable way) near the region where it ceases to be flat then the free boundary region associated to the flattens of u (*i.e.* the region where $u_h = u - h$ vanishes) may coincide with the boundary of the set where h is flat (see Theorem 6 for $f(t) = t^{\mathrm{q}}$, $\mathrm{q} < \mathrm{N}$). The estimates on the localization of the free boundary are optimal, in the class of nonlinearities $f(s)$ satisfying (1.6).

In Section 5, by means of a Strong Maximum Principle for u_h, we prove that the condition

$$\int_{0^+} \mathrm{F}(t)^{-\frac{1}{k+1}} = \infty \tag{1.7}$$

(or $k \leq \mathrm{q}$ for $f(t) = t^{\mathrm{q}}$) is a necessary condition for the non-existence of such free boundary (see Theorem 8 and Corollary 2). More precisely, we shall prove that under this condition the solution can not have any flat region. This can be regarded as an extension of [20] to the non divergence case (see also [7], [9] and [16]). As it was pointed out, the condition $k \leq \mathrm{q}$ implies non-degenerate ellipticity of problem (3.6) under very simple assumptions, such as $\varphi(x_0) > h(x_0)$ at some $x_0 \in \partial\Omega$ or $\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 h(x_0)\big)\big] > 0$ at some point $x_0 \in \Omega$ for path-connected open set Ω (see Corollary 2).

Finally, we note that all contributions coincide with the relative ones of [8] whenever $k = \mathrm{N}$.

2. Notations and other preliminaries comments

For any matrix $\mathbf{A} \in \mathcal{M}(\mathbb{R}, \mathrm{N} \times \mathrm{N})$, we consider the sum of the $k \times k$, $1 \le k \le \mathrm{N}$ principal minors, here denoted by

$$\mathcal{S}_k\big[\lambda(\mathbf{A})\big] = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le \mathrm{N}} \lambda_{i_1} \cdots \lambda_{i_k}, \tag{2.1}$$

where $\lambda(\mathbf{A}) = \big(\lambda_1, \ldots, \lambda_\mathrm{N}\big)$ are the eigenvalues of $\mathbf{A}$. Obviously, k is an integer number taking value in $[1, \mathrm{N}]$. The more popular examples are $\mathcal{S}_1[\lambda(\mathbf{A})] =$ trace of $\mathbf{A}$ and $\mathcal{S}_\mathrm{N}[\lambda(\mathbf{A})] = \det(\mathbf{A})$. The main important case appears when $\mathbf{A} = \mathrm{D}^2 u$ for some function $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$, where Ω is an open set of $\mathbb{R}^\mathrm{N}$, for which $\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u\big)\big]$ is called the k^{th} elementary symmetric function. If we define the cone

$$\Gamma_k = \left\{ \big(\lambda_1, \ldots, \lambda_\mathrm{N}\big) \in \mathbb{R}^\mathrm{N} : \sum_{1 \le i_1 < i_2 < \cdots < i_j \le \mathrm{N}} \lambda_{i_1} \cdots \lambda_{i_j} > 0, \ \forall j = 1, \ldots, k \right\}$$

the k–convex functions are introduced by the condition $\lambda\big(\mathrm{D}^2 u\big) \in \overline{\Gamma}_k$. Then, a function is 1–admissible if and only if it is sub–harmonic and a function N–admissible must be convex, because $\det \mathrm{D}^2 u > 0$ implies convexity by the Sylvester criterion. The expression $\mathcal{S}_k\big[\lambda(\mathbf{A})\big]$ are denoted alternatively as "principal invariants" of the tensor $\mathbf{A}$ (see [13, p.15]) as they are used in Continuum Mechanics. For instance the 2^{nd} elementary symmetric function plays a fundamental role in the study of Non–Newtonian fluids and Mooney–Rivlin materials (see [13, p. 174 and p. 192]).

Also we deduce that a k–admissible smooth function, for any $1 \le k \le \mathrm{N}$, is sub–harmonic because

$$\Gamma_\mathrm{N} \subset \cdots \subset \Gamma_k \subset \cdots \subset \Gamma_1.$$

Moreover the set of the k–admissible functions is a convex cone in $\mathcal{C}^2(\Omega)$. Since one proves that the matrix

$$\left\{ \frac{\partial \mathcal{S}_k[\lambda\big(\mathrm{D}^2 u\big)]}{\partial \mathrm{D}_{ij} u} : \ 1 \le i, j \le \mathrm{N} \right\} \tag{2.2}$$

is positive semi–definite if u is k–admissible, the k^{th} elementary symmetric operator $\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u\big)\big]$, in short the Hessian operator, is non–negative and degenerate elliptic on the convex cone of the k–admissible smooth functions (see [21]). Another main property useful to our reasoning: as function of $\mathrm{D}^2 u$

$$\Lambda_k\big(\mathrm{D}^2 u\big) = \big(\mathcal{S}_k[\lambda\big(\mathrm{D}^2 u\big)]\big)^{\frac{1}{k}}$$

is concave on the convex cone of the k–admissible functions. Finally we note that a compatibility geometric assumption must be required when one prescribes boundary values. In order to simplify it we only consider, for a while, smooth at the boundary k-admissible functions u vanishing on $\partial\Omega$. For any fixed point $x_0 \in \partial\Omega$, by suitable translations and rotations of coordinates if necessary, we may assume that x_0 is the origin and that locally $\partial\Omega$ is given by $x_{\mathrm{N}} = \Psi(x')$ such that $\mathbf{n} = (0, \dots, 0, 1)$ is the inner normal of $\partial\Omega$ at x_0, where $x' = (x_1, \dots, x_{\mathrm{N}-1})$. Then differentiating the boundary condition $u(x', \Psi(x')) = 0$ we get

$$\mathrm{D}_{ij}u(0) + \mathrm{D}_{\mathrm{N}}u(0)\mathrm{D}_{ij}\Psi(0) = 0.$$

Since u is sub–harmonic one has $\mathrm{D}_{\mathrm{N}}u(0) < 0$ whence

$$\frac{\partial \mathcal{S}_k[\lambda(\mathrm{D}^2 u)]}{\partial \mathrm{D}_{\mathrm{NN}}u} = |\mathrm{D}_{\mathrm{N}}u(0)|^{k-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq \mathrm{N}-1} \kappa_{i_1} \cdots \kappa_{i_k}$$

because the principal curvatures, $\kappa = (\kappa_1, \dots, \kappa_{\mathrm{N}-1})$, of $\partial\Omega$ at x_0 are the eigenvalues of $\mathrm{D}_{ij}\Psi(0)$. So that, as the matrix given by (2.2) is positive semi–definite we know one the following condition holds:

$$\sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq \mathrm{N}-1} \kappa_{i_1} \cdots \kappa_{i_k} \geq c_0 > 0 \quad \text{on } \partial\Omega \text{ for some constant } c_0.$$

This defines the $(k-1)$–convex domains that we will consider in this paper. Clearly, when $k = \mathrm{N}$ this geometric condition is equivalent to the usual convexity.

Many previous expositions on the nature of the k–admissible functions can be found in the literature (see for instance the survey [21] or [4]).

As it was proved by several methods [3, 4, 18, 21], there exists a k–admissible $\mathcal{C}^2$ solution of the general boundary value problems as

$$\begin{cases} \mathcal{S}_k\big[\lambda(\mathrm{D}^2 u)\big] = \mathrm{H}(\mathrm{D}u, u, x), & \text{on } \Omega, \\ u = \varphi, & \text{on } \partial\Omega, \end{cases} \tag{2.3}$$

under suitable assumptions on Ω, $\mathrm{H} > 0$ and φ. A main question arises: what happens if $\mathrm{H} \geq 0$. Now the degenerate ellipticity may occur and in general the regularity $\mathcal{C}^2$ of solutions can not be guaranteed. As it was pointed out in the Introduction, the so called "viscosity solutions" or the "generalized solutions" the adequate notions of solutions in our study. In fact, by means of reasoning as in [14], it can be proved that for $(k-1)$–convex domains Ω both notions coincide.

A short description of all that is as follows. First of all, the smooth k–admissibilty notion must be weaken. So, from now, by a k–admissible function u in Ω we mean an upper semi–continuous function in Ω such that $\{u = \infty\}$ has measure zero and

$$\int_\Omega u a_{ij} \mathrm{D}^2_{ij}\phi \leq 0 \quad \forall \phi \in \mathcal{C}_0^\infty(\Omega), \ \phi \geq 0,$$

for any matriz $\mathbf{A} = \{a_{ij}\}$ with eigenvalues in

$$\Gamma_k^* = \big\{\lambda^* \in \mathbb{R}^{\mathrm{N}} : \ \langle \lambda^*, \lambda\rangle \leq 0, \ \lambda \in \Gamma_k\big\}.$$

This implies that an upper semi–continuous function u is k–admissible is it is sub–harmonic with respect to the operator $\mathcal{L} = \sum a_{ij}\mathrm{D}^2_{ij}$ for any matriz $\mathbf{A} = \{a_{ij}\}$ with eigenvalues in Γ_k^*. Certainly, this non–smooth notion is consistent with the smooth k–admissible notion.

On the other hand, if $u \in \mathcal{C}^2(\Omega)$ is a non–negative k–admissible function the measure $\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2u\big)\big]dx$ has the important property that if $\{u_j\}_j \subset \mathcal{C}^2(\Omega)$ are smooth k–admissible functions which converge to a k–admissible function in Ω everywhere then $\big\{\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2u_j\big)\big]dx\big\}_j$ converge weakly to a measure μ. With this property one proves

Theorem 1 ([21]) *For any k–admissible function u, there exists a Radon measure $\mu_k[u]$ such that:*

1. $\mu_k[u] = \mathcal{S}_k\big[\lambda\big(\mathrm{D}^2u\big)\big]dx$ *if* $u \in \mathcal{C}^2(\Omega)$,
2. *if $\{u_j\}_j$ are k–admissible functions which converge to a k–admissible function u a.e. then $\big\{\mu_k[u_j]\big\}_j \to \mu[u]$ weakly as measure.* □

Then we arrive to

Definition 2.1 *A k–admissible function u on Ω is a "generalized solution" of*

$$\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2u\big)\big] = \mathrm{H}(\mathrm{D}u, u, x), \quad \textit{on } \Omega$$

if

$$\mu[u](\mathrm{E}) = \int_{\mathrm{E}} \mathrm{H}(\mathrm{D}u, u, x)dx$$

for any Borel set $\mathrm{E} \subset \Omega$.

The continuity property of u on $\overline{\Omega}$ is compatible with the usual realization of the Dirichlet boundary condition $u = \varphi$. Here we may considered the weaker assumption $\mathrm{H} \geq 0$ which can not be removed. Certainly, the definition can be extended to locally k–admissible functions u on Ω, for which u can be constant on some subset of Ω. This notion of generalized solution is specific of the Hessian operator, but other notion of solutions are available as it happens with other type of fully nonlinear equations in non divergence form. The most usual is the so called "viscosity solution" introduced by M.G. Crandall and P.L. Lions (see [6]):

Definition 2.2 *A k–admissible function u on Ω is a viscosity solution of the inequality*

$$\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2u\big)\big] \geq \mathrm{H}(\mathrm{D}u, u, x) \quad \textit{in } \Omega \qquad \textit{(viscosity sub–solution)}$$

if for every smooth k–admissible function Φ on Ω for which

$$(u-\Phi)(x_0) \geq (u-\Phi)(x) \quad \textit{locally at } x_0 \in \Omega$$

one has

$$\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2\Phi(x_0)\big)\big] \geq \mathrm{H}\big(\mathrm{D}\Phi(x_0), u(x_0), x_0\big).$$

Analogously, one defines the viscosity solution of the reverse inequality

$$\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u\big)\big] \leq \mathrm{H}(\mathrm{D}u, u, x) \quad \textit{in } \Omega \qquad \textit{(viscosity super–solution)}$$

as a k–admissible function u on Ω such that for every smooth k–admissible function Φ on Ω for which

$$(u-\Phi)(x_0) \leq (u-\Phi)(x) \quad \textit{locally at } x_0 \in \Omega$$

one has

$$\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2\Phi(x_0)\big)\big] \leq \mathrm{H}\big(\mathrm{D}\Phi(x_0), u(x_0), (x_0)\big).$$

Finally, when both properties hold we arrive to the notion of viscosity solution of

$$\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u\big)\big] = \mathrm{H}(\mathrm{D}u, u, x) \quad \textit{in } \Omega.$$

Note that the k–admissible condition on u and Φ are extra assumptions with respect to the usual notion of viscosity solution (see [6]). This is needed here because the Hessian operator is only degenerate elliptic on this class of functions. In fact, the k–admissible condition on the test function Φ is only required for the correct definition of super–solutions in the viscosity sense, because if $u-\Phi$ attains a local maximum at $x_0 \in \Omega$ for a k–admissible function u on Ω and $\Phi \in \mathcal{C}^2(\Omega)$, reasoning as in [14], one can deduce

$$\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2\Phi(x_0)\big)\big] \geq 0.$$

Reasoning again as in [14], it is possible to prove the equivalence

u is a generalized solution of (2.3) *if and only if u is a viscosity solution of* (2.3),

provided that Ω is a $(k-1)$–convex domain and $\mathrm{H} \in \mathcal{C}(\mathbb{R}^{\mathrm{N}} \times \mathbb{R} \times \Omega)$.

As an illustrative result on the complementary regularity, one proves that $\mathrm{H}(\mathrm{D}u, u, x) \in \mathrm{L}^p(\Omega)$, $p > \frac{\mathrm{N}}{2k}$, implies that u is Hölder continuous, provided $k \leq \frac{\mathrm{N}}{2}$ (see [21, Corollary 9.1]).

3. Weak Maximum Principle

In this section we obtain some comparison and existence results for the equation (1.2). They will show that the nature of the viscosity solution is an intrinsic property associated with the Maximum Principle.

Theorem 2 (Weak Maximum Principle I) *Let $u, v \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$ where u is k–admissible in Ω. Suppose*

$$\mathcal{F}(\mathrm{D}^2 u, \mathrm{D}u, u) \leq 0 \leq \mathcal{F}(\mathrm{D}^2 v, \mathrm{D}v, v) \quad \text{in } \Omega. \tag{3.1}$$

Then

$$(u - v)(x) \leq \sup_{\partial\Omega} [u - v]_+, \quad x \in \Omega.$$

In particular,

$$|u - v|(x) \leq \sup_{\partial\Omega} |u - v|, \quad x \in \Omega.$$

whenever the equalities hold in (3.1).

Proof. By continuity there exists $x_0 \in \overline{\Omega}$ where $[u - v]_+$ attains the maximum value on $\overline{\Omega}$. We claim that $[u - v]_+(x_0) = 0$, whence the result follows. Indeed, if $x_0 \in \Omega$ and $[u - v]_+(x_0) > 0$, the matrix $\mathrm{D}^2(v - u)(x_0)$ is positive semidefinite. In particular, the function $v - u$ is k–admissible at x_0. Consider the function $\Lambda_k(\mathbf{A}) \doteq \big(\mathcal{S}_k[\lambda(\mathbf{A})]\big)^{\frac{1}{k}}$ which is homogeneous of degree 1 and concave on the convex cone of the set of matrices having eigenvalues in Γ_k. Since the convexity of this set of matrices implies that the sum of two k–admissible functions is also k-admissible, the function $v = (v - u) + u$ is k–admissible at x_0. Then

$$\Lambda_k[\mathrm{D}^2 v](x_0) = 2^k \Lambda_k \left[\frac{1}{2}\mathrm{D}^2 v\right](x_0) \geq \Lambda_k[\mathrm{D}^2(v - u)](x_0) + \Lambda_k[\mathrm{D}^2 u](x_0) \geq \Lambda_k[\mathrm{D}^2 u](x_0)$$

leads to the contradiction

$$\begin{aligned} 0 &\leq \Lambda_k[\mathrm{D}^2 v](x_0) - \Lambda_k[\mathrm{D}^2 u](x_0) \\ &\leq \big(g\big(|\mathrm{D}v(x_0)|\big) f\big(v(x_0)\big) - h(x_0)\big)^{\frac{1}{k}} - \big(g\big(|\mathrm{D}u(x_0)|\big) f\big(u(x_0)\big) - h(x_0)\big)^{\frac{1}{k}} < 0. \end{aligned}$$

□

Remark 3.1 We note that the monotonicity on u on the zeroth order terms, $f(u-h)$, is the only assumption required on the structure of the equation and that our argument is strongly based on the notion of viscosity solution. An analogous estimate holds by changing the roles of u_1 and u_2 (but then we do not require the $\mathcal{C}^2$ function u_1 to be k–admissible). Note also that we did not assume any convexity condition on the domain Ω. When Ω is $(k - 1)$–convex these results can be extended to the class of the generalized solutions through the mentioned equivalence between such solutions and the viscosity solutions. □

A very simple (and important fact) was used in our precedent arguments: if $u_1 \in \mathcal{C}^2$ and $u_2 - u_1 \in \mathcal{C}^2$ are k–admissible functions on a ball $\mathbf{B}$ then

$$\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u_2\big)\big] \geq \mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u_1\big)\big] \quad \text{in } \mathbf{B}.$$

This simple inequality can be extended to the case where u_1 and $u_2 - u_1$ are k–admissible functions on a ball $\mathbf{B}$, with $u_1 = u_2$ on $\partial\mathbf{B}$, by the "monotonicity formula"

$$\mu[u_2](\mathbf{B}) \geq \mu[u_1](\mathbf{B}) \tag{3.2}$$

(see [21]). In this way, the Weak Maximum Principle can be extended to the class of generalized solutions.

Theorem 3.1 (Weak Maximum Principle II) *Let $h_1, h_2 \in \mathcal{C}(\overline{\Omega})$. Let $u_1, u_2 \in \mathcal{C}(\overline{\Omega})$ where u_1 is locally k–admissible in Ω. Suppose*

$$-\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u_1\big)\big] + g\big(|\mathrm{D}u_1|\big) f\big(u_1 - h_1\big) \leq -\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u_2\big)\big] + g\big(|\mathrm{D}u_2|\big) f\big(u_2 - h_2\big) \quad \textit{in } \Omega \tag{3.3}$$

in the generalized solutions sense. Then

$$(u_1 - u_2)(x) \leq \sup_{\partial\Omega}[u_1 - u_2]_+ + \sup_{\Omega}[h_1 - h_2]_+, \quad x \in \overline{\Omega}. \tag{3.4}$$

In particular,

$$|u_1 - u_2|(x) \leq \sup_{\partial\Omega}|u_1 - u_2| + \sup_{\Omega}|h_1 - h_2|, \quad x \in \overline{\Omega}, \tag{3.5}$$

whenever the equality holds in (3.3).

Proof. As above, we only consider the case where the maximum of $[u_1 - u_2]_+$ on $\overline{\Omega}$ is achieved at some $x_0 \in \Omega$ with $[u_1 - u_2]_+(x_0) > 0$. Therefore, $(u_1 - u_2\big)(x) > 0$ and convex in a ball $\mathbf{B}_{\mathrm{R}}(x_0)$, for R small. Let $\Omega^+ = \{u_1 > u_2\} \supseteq \mathbf{B}_{\mathrm{R}}(x_0)$. We construct $\widehat{u}_1(x) = u_1(x) + \gamma\big(|x - x_0|^2 - \mathrm{M}^2\big) - \delta$, where $\mathrm{M} > 0$ is large and $\gamma, \delta > 0$ such that $\widehat{u}_1 < u_1$ on $\partial\Omega^+$ and the set $\Omega^+_{\gamma,\delta} = \{\widehat{u}_1 > u_2\}$ is compactly contained in Ω and contains $\mathbf{B}_\varepsilon(x_0)$ for some ε small. By choosing γ, δ properly, we can assume that the diameter of $\Omega^+_{\gamma,\delta}$ is small so that u_1 and therefore $u_2 = (u_2 - u_1) + u_1$ are convex in it. Then (3.2) implies

$$\begin{aligned} 0 < (\gamma\varepsilon)^{\mathrm{N}}|\mathbf{B}_1(0)| &\leq \mu[u_2]\big(\mathbf{B}_\varepsilon(x_0)\big) - \mu[u_1]\big(\mathbf{B}_\varepsilon(x_0)\big) \\ &\leq \int_{\mathbf{B}_\varepsilon(x_0)} \big[g\big(|\mathrm{D}u_2|\big) f\big(u_2 - h_2\big) - g\big(|\mathrm{D}u_1|\big) f\big(u_1 - h_1\big)\big] dx. \end{aligned}$$

Since $g\big(|\mathrm{D}u_1(x_0)|\big) = g\big(|\mathrm{D}u_2(x_0)|\big) > 0$ (see Remark 3.2 below), by letting $\varepsilon \to 0$, the Lebesgue differentiation theorem implies

$$0 \leq g\big(|\mathrm{D}u_2(x_0)|\big) f\big(u_2(x_0) - h_2(x_0)\big) - g\big(|\mathrm{D}u_1(x_0)|\big) f\big(u_1(x_0) - h_1(x_0)\big),$$

whence

$$\big(u_1 - u_2\big)(x_0) < \big(h_1 - h_2\big)(x_0) \leq \sup_{\partial\Omega}\big[u_1 - u_2\big]_+ + \sup_{\Omega}\big[h_1 - h_2\big]_+$$

and the estimate holds. $\square$

Remark 3.2 The above proof requires a simple fact: any convex function ψ in a convex open set $\mathcal{O} \subset \mathbb{R}^N$ achieving a local interior maximum at some $z_0 \in \mathcal{O}$ verifies $\mathrm{D}\psi(z_0) = \mathbf{0}$. Indeed, for any $\mathbf{p} \in \partial\psi(z_0)$ (the sub–differential set of ψ at z_0) one has

$$\psi(x) \geq \psi(z_0) + \langle \mathbf{p}, x - z_0 \rangle \geq \psi(x) + \langle \mathbf{p}, x - z_0 \rangle \quad \text{with } x \text{ near } z_0,$$

and

$$0 \geq \langle \mathbf{p}, x - z_0 \rangle.$$

Then if $\tau > 0$ is small enough we may choose $x - z_0 = \tau\mathbf{p} \in \mathcal{O}$ and to deduce

$$0 \leq \tau|\mathbf{p}|^2 \leq 0.$$

□

A first consequence of the general theory for (1.2) and the Weak Maximum Principle is the following existence result

Theorem 3.2 *Suppose that Ω is $(k-1)$–convex. Let $\varphi \in \mathcal{C}(\partial\Omega)$ and assume the compatibility condition* (1.4)

$$h \text{ is locally } k\text{-admissible on } \overline{\Omega} \text{ and } h \leq u \text{ on } \partial\Omega.$$

Then there exists a unique locally k–admissible function verifying

$$\begin{cases} \mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u\big)\big] = g\big(|\mathrm{D}u|\big) f(u-h) & \text{in } \Omega, \\ u = \varphi & \text{on } \partial\Omega, \end{cases} \tag{3.6}$$

in the generalized sense. In fact, one verifies

$$h(x) \leq u(x) \leq \mathrm{U}_\varphi(x), \quad x \in \overline{\Omega}, \tag{3.7}$$

where U_φ is the unique harmonic function in Ω such that $\mathrm{U}_\varphi = \varphi$ on $\partial\Omega$.

Proof. First we consider the generalized solution of the problem

$$\begin{cases} -\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u\big)\big] + g\big(|\mathrm{D}u|\big)\big[f(u-h)\big]_+ = 0 & \text{in } \Omega, \\ u = \varphi & \text{on } \partial\Omega. \end{cases}$$

Since $\mathrm{H}\big(\mathrm{D}u, u, x\big) = g\big(|\mathrm{D}u|\big)\big[f(u-h)\big]_+ \geq 0$ we can apply well known results in the literature. In particular, from [21], it follows the existence and uniqueness of the solution u. The second point is to note that, by construction, the locally k–admissible function h verifies

$$-\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 h\big)\big] + g\big(|\mathrm{D}u|\big)\big[f(h-h)\big]_+ \leq 0 \quad \text{in } \Omega.$$

Therefore, by the Weak Maximum Principle and the assumption $h \leq \varphi$ on $\partial\Omega$ we get that

$$h \leq u \quad \text{in } \Omega,$$

whence

$$\big[f(u-h)\big]_+ = f(u-h)$$

which proves that u solves (3.6). The uniqueness also follows from the Weak Maximum Principle. Finally, since u is locally k–admissible it is also sub–harmonic in Ω and so the estimate

$$h(x) \leq u(x) \leq \mathrm{U}_\varphi(x), \quad x \in \overline{\Omega}$$

holds by the weak maximum principe for harmonic functions. □

Remark 3.3 i) As it was pointed out in the Introduction, no sign assumption on h is required in Theorem 3.2. The simple structural assumption (1.4) implies that $h \leq u$ on $\overline{\Omega}$ and therefore the ellipticity, eventually degenerate, of the equation holds. Thus, the ellipticity holds once h behaves as a lower "obstacle" for the solution u. We note that these compatibility conditions are not a priori required in the Weak Maximum Principle because there we are working with functions whose existence is a priori assumed.
ii) Since u is locally k–admissible on $\overline{\Omega}$, we can prove

$$\sup_\Omega |\mathrm{D}u| = \sup_{\partial\Omega} |\mathrm{D}u|,$$

and then inequality (3.7) gives a priori bounds on $|\mathrm{D}u|$ on $\overline{\Omega}$, provided $h = \varphi$ on $\partial\Omega$ and $\mathrm{D}h$ is defined on $\partial\Omega$. The proof of a second derivative estimate is based on the inequality

$$\operatorname*{ess\,sup}_\Omega |\mathrm{D}^2 u| \leq \mathrm{C}\left(1 + \sup_{\partial\Omega} |\mathrm{D}^2 u|\right) \tag{3.8}$$

for some constant C independent on u. It will be the object of a separated article. □

In Section 5 we shall prove a kind of Strong Maximum Principle which under suitable assumptions will avoid the appearance of the mentioned free boundary.

4. Flat regions

In this section we focus our attention on a lower "obstacle" function h locally k–admissible on $\overline{\Omega}$ which is locally flat. We define

$$\mathrm{Flat}(h) = \bigcup_\alpha \mathrm{Flat}_\alpha(h)$$

where

$$\mathrm{Flat}_\alpha(h) = \{x \in \overline{\Omega} : h(x) = \langle \mathbf{p}_\alpha, x\rangle + a_\alpha,\ \text{for some } \mathbf{p}_\alpha \in \mathbb{R}^{\mathrm{N}} \text{ and } a_\alpha \in \mathbb{R}\}. \tag{4.1}$$

Since

$$u(y) - \big(\langle \mathbf{p}_\alpha, y\rangle + a_\alpha\big) \geq u(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big) + \langle \mathbf{p} - \mathbf{p}_\alpha, y - x\rangle,$$

thus

$$\mathbf{p} \in \partial u(x) \quad \Leftrightarrow \quad \mathbf{p} - \mathbf{p}_\alpha \in \partial\left(u(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big)\right),$$

and the equation (1.2) becomes

$$-\mathcal{S}_k[\lambda(\mathrm{D}^2 u_\alpha)] = \eta g\big(|\mathrm{D}u|\big) f\big(u_\alpha\big), \quad x \in \mathrm{Flat}_\alpha(h), \tag{4.2}$$

for $u_\alpha \doteq u - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big)$. Remember that $u_\alpha \geq 0$ in an open set $\mathcal{O} \subseteq \Omega$, if $u_h \geq 0$ on $\partial\mathcal{O}$. Assumption

$$g(|\mathbf{p}|) \geq 1 \tag{4.3}$$

leads us to study the auxiliar boundary problem

$$\begin{cases} \mathcal{S}_k[\lambda(\mathrm{D}^2\mathrm{U})] = f(\mathrm{U}) & \text{in } \mathbf{B}_{\mathrm{R}}(0), \\ \mathrm{U} \equiv \mathrm{M} > 0 & \text{on } \partial\mathbf{B}_{\mathrm{R}}(0), \end{cases} \tag{4.4}$$

for any $\mathrm{M} > 0$. From the uniqueness of solutions, it follows that U is radially symmetric, because by rotating it we would find other solutions. Moreover, by the comparison results U is nonnegative. Therefore, the solution U is governed by a non–negative radial profile function $\mathrm{U}(x) = \widehat{\mathrm{U}}(|x|)$ for which some straightforward computations leads to

$$\begin{aligned} \mathcal{S}_k[\lambda(\mathrm{D}^2\mathrm{U})](r) &= \mathrm{C}_{\mathrm{N}-1,k-1}\widehat{\mathrm{U}}''(r)\left(\frac{\widehat{\mathrm{U}}'(r)}{r}\right)^{k-1} + \mathrm{C}_{\mathrm{N}-1,k}\left(\frac{\widehat{\mathrm{U}}'(r)}{r}\right)^{k} \\ &= \mathrm{C}_{\mathrm{N}-1,k-1} r^{1-\mathrm{N}}\left[\frac{r^{\mathrm{N}-k}}{k}\big(\widehat{\mathrm{U}}'(r)\big)^k\right]', \end{aligned} \tag{4.5}$$

where we use the notation

$$\mathrm{C}_{m,n} = \binom{m}{n} = \frac{m!}{(m-n)!n!}, \quad 0 \leq n \leq m.$$

We summarize the well known properties:

$$\begin{cases} \mathrm{C}_{m,0} = \mathrm{C}_{m,m} & \text{(initial/boundary values)}, \\ \mathrm{C}_{m,n} = \mathrm{C}_{m,m-n} & \text{(symmetry)}, \\ \mathrm{C}_{m,k} + \mathrm{C}_{m+k+1} = \mathrm{C}_{m+1,k+1} & \text{(recursive Pascal rule)}. \end{cases}$$

Remark 4.1 For $\mathrm{N} = 1$, the equation (4.5) becomes the semi linear ODE

$$\widehat{\mathrm{U}}''(r) = \lambda f\big(\widehat{\mathrm{U}}\big)$$

studied in [9]. Notice that for $\mathrm{N} > 1$ the k–radial Hessian operator *is not* exactly the radial p-Laplacian operator with $p = k + 1$, although there is a great resemblance among them. □

We start this section by considering the initial value problem

$$\begin{cases} \mathrm{C}_{\mathrm{N}-1,k-1} r^{1-\mathrm{N}} \left[\dfrac{r^{\mathrm{N}-k}}{k} \big(\mathrm{U}'(r)\big)^k \right]' = \eta f\big(\mathrm{U}(r)\big), \quad \eta > 0, \\ \mathrm{U}(0) = \mathrm{U}'(0) = 0. \end{cases} \tag{4.6}$$

Obviously, $\mathrm{U}(r) \equiv 0$ is always a solution, but we are interested in the existence of nontrivial and non–negative solutions. The general reasoning in this section assumes the existence of an increasing function $\mathbb{U} : [0, \mathrm{R}_{\mathbb{U}}[\to \overline{\mathbb{R}}_+$ solving

$$\begin{cases} \mathrm{C}_{\mathrm{N}-1,k-1} r^{1-\mathrm{N}} \left[\dfrac{r^{\mathrm{N}-k}}{k} \big(\mathbb{U}'(r)\big)^k \right]' = \eta_{\mathbb{U}} f\big(\mathbb{U}(r)\big), \quad 0 < r < \mathrm{R}_{\mathbb{U}}, \\ \mathbb{U}(0) = \mathbb{U}'(0) = 0, \end{cases} \tag{4.7}$$

for some $\eta_{\mathbb{U}} > 0$ and $0 < \mathrm{R}_{\mathbb{U}} \leq \infty$. We shall return to this assumption later.

By scaling by $\mathrm{A} > 0$, (4.7) becomes

$$\begin{cases} -\mathrm{C}_{\mathrm{N}-1,k-1} r^{1-\mathrm{N}} \left[\dfrac{r^{\mathrm{N}-k}}{k} \big(\widehat{\mathbb{U}}'(\mathrm{A}r)\big)^k \right]' + \eta f\big(\mathbb{U}(\mathrm{A}r)\big) = \big[\eta - \eta_{\mathbb{U}} \mathrm{A}^{2k}\big] f\big(\mathbb{U}(\mathrm{A}r)\big) \\ \mathbb{U}(0) = \mathbb{U}'(0) = 0, \end{cases} \tag{4.8}$$

$0 < r < \dfrac{\mathrm{R}_{\mathbb{U}}}{\mathrm{A}}$, whence it follows

1. if $\mathrm{A} < \left(\dfrac{\eta}{\eta_{\mathbb{U}}}\right)^{\frac{1}{2k}}$ the function $\mathbb{U}(\mathrm{A}r)$ is a super–solution of the equation (4.6),
2. if $\mathrm{A} = \left(\dfrac{\eta}{\eta_{\mathbb{U}}}\right)^{\frac{1}{2k}}$ the function $\mathbb{U}(\mathrm{A}r)$ is the solution of the equation (4.6),
3. if $\mathrm{A} > \left(\dfrac{\eta}{\eta_{\mathbb{U}}}\right)^{\frac{1}{2k}}$ the function $\mathbb{U}(\mathrm{A}r)$ is a sub–solution of the equation (4.6).

Moreover, the function

$$v_\tau(x) \doteq \mathbb{U}\left(\left(\frac{\eta}{\eta_{\mathbb{U}}}\right)^{\frac{1}{2k}} [|x| - \tau]_+ \right), \quad x \in \mathbf{B}_{\tau + \mathrm{R}_{\mathbb{U},\eta}}(0),\ \mathrm{R}_{\mathbb{U},\eta} = \mathrm{R}_{\mathbb{U}} \left(\frac{\eta_{\mathbb{U}}}{\eta}\right)^{\frac{1}{2k}} \tag{4.9}$$

solves

$$-\mathcal{S}_k[\lambda(\mathrm{D}^2 v_\tau(x)] + \eta f\big(v_\tau(x)\big) = 0, \quad x \in \mathbf{B}_{\tau+\mathrm{R}_{\mathbb{U},\eta}}(0).$$

Furthermore, it verifies

$$v_\tau(x) = \mathrm{M}, \quad |x| = \mathrm{R} < \tau + \mathrm{R}_{\mathbb{U},\eta}$$

once we take

$$\tau = \mathrm{R} - \left(\frac{\eta_{\mathbb{U}}}{\eta}\right)^{\frac{1}{2k}} \mathbb{U}^{-1}(\mathrm{M}) = \left[\eta_*^{-\frac{1}{2k}} - \eta^{-\frac{1}{2k}}\right] \mathbb{U}^{-1}(\mathrm{M})\eta_{\mathbb{U}}^{\frac{1}{2k}}$$

with

$$\eta \geq \eta_* \doteq \eta_{\mathbb{U}} \left(\frac{1}{\mathrm{R}}\mathbb{U}^{-1}(\mathrm{M})\right)^{2k}. \tag{4.10}$$

Now for a solution of (1.2) we may localize a core of the flat region $\mathrm{Flat}(u)$ inside the flat subregion $\mathrm{Flat}_\alpha(h)$ of the given "obstacle".

Theorem 3 *Let h be locally k–admissible on $\overline{\Omega}$. Let us assume that there exists $\mathbf{B}_{\mathrm{R}}(x_0) \subset \mathrm{Flat}_\alpha(h)$ with*

$$0 \leq u(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big) \leq \mathrm{M} \leq \max_{\overline{\Omega}}(u-h), \quad x \in \partial\mathbf{B}_{\mathrm{R}}(x_0), \tag{4.11}$$

where u is a generalized solution of (1.2)*, for some $\mathrm{M} > 0$. Then, if* (4.7) *holds and*

$$\eta \geq \eta_* \doteq \eta_{\mathbb{U}} \left(\frac{1}{\mathrm{R}}\mathbb{U}^{-1}(\mathrm{M})\right)^{2k},$$

one verifies

$$0 \leq u(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big) \leq \mathbb{U}\left(\left(\frac{\eta}{\eta_{\mathbb{U}}}\right)^{\frac{1}{2k}} [|x| - \tau]_+\right), \quad x \in \mathbf{B}_{\mathrm{R}}(x_0), \tag{4.12}$$

where

$$\tau = \left[\eta_*^{-\frac{1}{2k}} - \eta^{-\frac{1}{2k}}\right] \mathbb{U}^{-1}(\mathrm{M})\eta_{\mathbb{U}}^{\frac{1}{2k}}, \tag{4.13}$$

once we assume that $\mathrm{R} < \tau + \mathrm{R}_{\mathbb{U},\eta}$ and

$$\left(\frac{\eta_{\mathbb{U}}}{\eta}\right)^{\frac{1}{2k}} \mathbb{U}^{-1}(\mathrm{M}) < \mathrm{R} \leq \mathrm{dist}(x_0, \partial\Omega). \tag{4.14}$$

In particular, the function u is flat on $\overline{\mathbf{B}}_\tau(x_0)$. More precisely,

$$u(x) = \langle \mathbf{p}_\alpha, x\rangle + a_\alpha \quad \textit{for any } x \in \overline{\mathbf{B}}_\tau(x_0).$$

Proof. The result is a direct consequence of previous arguments. Indeed, for simplicity we can assume $x_0 = 0$. Since $g(|\mathbf{p}|) \geq 1$, by the comparison results we get that

$$0 \leq u_\alpha(x) \leq v_\tau(x), \quad x \in \mathbf{B}_{\mathrm{R}}(0)$$

(see (4.2) and (4.9)) and so the conclusions hold. □

Remark 1 We have proved that under the above assumptions the flat region of u is a non–empty set. Obviously, $\mathrm{Flat}(h) \subset \mathrm{Flat}(u)$ whenever (4.11) fails, even if (4.7) holds. □

Remark 2 We point out that the above result applies to the case in which $\varphi \equiv 1$ and $h \equiv 0$ (the so called "dead core" problem) as well as to cases in which u is flat only near $\partial\Omega$ (take for instance, $h(x) = \langle \mathbf{p}_\alpha, x\rangle + a_\alpha$ in Ω and $\varphi \equiv h$ on $\partial\Omega$). □

In order to study the assumption (4.7) we note that the function $\mathbb{U}(r)$ satisfies the inequality

$$\mathrm{C}_{\mathrm{N}-1,k-1}\mathbb{U}''(r)\left(\frac{\mathbb{U}'(r)}{r}\right)^{k-1} \leq \eta_{\mathbb{U}} f\big(\mathbb{U}(r)\big), \quad 0 < r < \mathrm{R}_{\mathbb{U}}, \tag{4.15}$$

(see (4.5)) whence

$$\Big(\big(\mathbb{U}'(r)\big)^{k+1}\Big)' \leq \eta_{\mathbb{U}}(k+1)\mathrm{C}^{-1}_{\mathrm{N}-1,k-1} r^{k-1}\left(\mathrm{F}\big(\mathbb{U}(r)\big)\right)', \quad 0 < r < \mathrm{R}_{\mathbb{U}} \quad \mathrm{F}' = f,$$

and

$$\big(\mathbb{U}'(r)\big)^{k+1} \leq \eta_{\mathbb{U}}(k+1)\mathrm{C}^{-1}_{\mathrm{N}-1,k-1} r^{k-1}\mathrm{F}\big(\mathbb{U}(r)\big), \quad 0 < r < \mathrm{R}_{\mathbb{U}}.$$

So, we deduce that (4.7) requires

$$\int_0^{\mathbb{U}(r)} \frac{ds}{\mathrm{F}(s)^{\frac{1}{k+1}}} = \int_0^r \frac{\mathbb{U}'(s)ds}{\mathrm{F}\big(\mathbb{U}(s)\big)^{\frac{1}{k+1}}} \leq \Big(\eta_{\mathbb{U}}(k+1)\mathrm{C}^{-1}_{\mathrm{N}-1,k-1}\Big)^{\frac{1}{k+1}} \frac{\mathrm{N}+1}{2k} r^{\frac{2k}{k+1}}$$

for $0 < r < \mathrm{R}_{\mathbb{U}}$. Therefore (1.6) is a necessary condition in order to (4.7) holds.

The reasoning in proving that (1.6) is a sufficient condition for the assumption (4.7) is very laborious and follows from some adaptations of the results of [9]. Here we only construct a function verifying a similar property for (4.15) useful to our interest

Theorem 4 *Assume* (1.6). *Then the function* $\phi(r)$ *given implicity by*

$$\int_0^{\phi(r)} \mathrm{F}(s)^{-\frac{1}{k+1}} ds = r^{\frac{2k-1}{k}}, \quad 0 \leq r \tag{4.16}$$

satisfies, the property

$$\begin{cases} \mathrm{C}_{\mathrm{N}-1,k-1} r^{1-\mathrm{N}}\left[\dfrac{r^{\mathrm{N}-k}}{k}\big(\phi'(r)\big)^k\right]' \leq \eta_{\phi,\widehat{\mathrm{R}}} f\big(\phi(r)\big), \quad 0 < r < \widehat{\mathrm{R}}, \\ \phi(0) = \phi'(0) = 0, \end{cases} \tag{4.17}$$

where

$$\left\{\begin{array}{l} \widehat{\mathrm{R}} < \displaystyle\int_0^\infty \mathrm{F}(s)^{-\frac{1}{k+1}} ds \leq +\infty, \\ \eta_{\phi,\widehat{\mathrm{R}}} = \mathrm{C}_{\mathrm{N}-1,k-1} \left(\dfrac{2k-1}{k}\right)^{k+1} \dfrac{\mathrm{N}}{k+1} \widehat{\mathrm{R}}^{\frac{k-1}{k}}. \end{array}\right. \tag{4.18}$$

Proof. Since the function

$$\psi(t) = \int_0^t \mathrm{F}(s)^{-\frac{1}{k+1}} ds, \quad t \geq 0,$$

is increasing from $\overline{\mathbb{R}}_+$ to $[0, \psi(\infty)[$ and $\psi(0) = 0$, we may consider the function $\phi(r)$ given implicitly by

$$\int_0^{\phi(r)} \mathrm{F}(s)^{-\frac{1}{k+1}} ds = r^a, \quad 0 \leq r < \psi(\infty) \leq +\infty,$$

where a is a positive constant to be chosen. Then

$$\phi'(r) = a\mathrm{F}\big(\phi(r)\big)^{\frac{1}{k+1}} r^{a-1},$$

and

$$\begin{aligned} r^{1-\mathrm{N}} \left[\frac{r^{\mathrm{N}-k}}{k} \big(\phi'(r)\big)^k\right]' &= \frac{a^k r^{1-\mathrm{N}}}{k} \left[r^{\mathrm{N}-2k+ka} \mathrm{F}\big(\phi(r)\big)^{\frac{k}{k+1}}\right]' \\ &= \frac{a^k r^{1-2k+ka}}{k} \left[\frac{\mathrm{N}-2k+ka}{r} \mathrm{F}\big(\phi(r)\big)^{\frac{k}{k+1}} + \frac{ak}{k+1} r^{a-1} f\big(\phi(r)\big)\right] \end{aligned}$$

hold. Next, we choose

$$a = \frac{2k-1}{k},$$

and $\Phi(r) = \big(\mathrm{F}\big(\phi(r)\big)\big)^{\frac{k}{k+1}}$. Since $\Phi(0) = 0$ and

$$\Phi'(r) = \frac{2k-1}{k+1} f\big(\phi(r)\big) r^{\frac{k-1}{k}}$$

is increasing, the convexity inequality

$$\Phi(r) \leq \Phi'(r) r$$

gives

$$\mathrm{C}_{\mathrm{N}-1,k-1} r^{1-\mathrm{N}} \left[\frac{r^{\mathrm{N}-k}}{k} \big(\phi'(r)\big)^k\right]' \leq \mathrm{C}_{\mathrm{N}-1,k-1} \left(\frac{2k-1}{k}\right)^{k+1} \frac{\mathrm{N}}{k+1} r^{\frac{k-1}{k}} f\big(\phi(r)\big).$$

Finally, since $a \geq 1$ one has $\phi(0) = \phi'(0) = 0$. $\square$

So that, fixed $\widehat{\mathrm{R}} < \psi(\infty)$ we have

$$\mathrm{C}_{\mathrm{N}-1,k-1} r^{1-\mathrm{N}} \left[\frac{r^{\mathrm{N}-k}}{k} \big(\phi'(\mathrm{A}r)\big)^k \right]' + \eta f\big(\phi(\mathrm{A}r)\big) \geq \left[\eta - \eta_{\phi,\widehat{\mathrm{R}}} \mathrm{A}^{2k} \right] f\big(\phi(\mathrm{A}r)\big), \quad (4.19)$$

for $0 < r < \widehat{\mathrm{R}}$ and $\phi(0) = \phi'(0) = 0$ (see (4.8)), whence the function

$$v_\tau(x) \doteq \phi \left(\left(\frac{\eta}{\eta_{\phi,\widehat{\mathrm{R}}}} \right)^{\frac{1}{2k}} [|x| - \tau]_+ \right), \quad x \in \mathbf{B}_{\tau + \mathrm{R}_{\phi,\eta}}(0), \quad (4.20)$$

solves

$$-\mathcal{S}_k[\lambda(\mathrm{D}^2 v_\tau(x)] + \eta f\big(v_\tau(x)\big) \geq 0, \quad x \in \mathbf{B}_{\tau + \mathrm{R}_{\mathbb{U}_{\phi,\eta,\widehat{\mathrm{R}}}}}(0),$$

for

$$\mathrm{R}_{\phi,\eta,\widehat{\mathrm{R}}} = \left(\frac{\eta_{\phi,\widehat{\mathrm{R}}}}{\eta} \right)^{\frac{1}{2k}} \widehat{\mathrm{R}}.$$

The reasonings of Theorem 3 apply and enable us to localize again a core of the flat region $\mathrm{Flat}(u)$ but now using the function ϕ given by (4.16) instead to use the function $\mathbb{U}$ given by (4.7).

Corollary 1 *Let h be locally k–admissible on $\overline{\Omega}$. Let us assume that there exists $\mathbf{B}_{\mathrm{R}}(x_0) \subset \mathrm{Flat}_\alpha(h)$ with*

$$0 \leq u(x) - \big(\langle \mathbf{p}_\alpha, x \rangle + a_\alpha\big) \leq \mathrm{M} \leq \max_{\overline{\Omega}}(u - h), \quad x \in \partial \mathbf{B}_{\mathrm{R}}(x_0), \quad (4.21)$$

where u is a generalized solution of (1.2)*, for some* $\mathrm{M} > 0$*. Then, if* (1.6) *holds and*

$$\eta \geq \eta_* \doteq \eta_{\phi,\widehat{\mathrm{R}}} \left(\frac{1}{\mathrm{R}} \phi^{-1}(\mathrm{M}) \right)^{2\mathrm{N}},$$

one verifies

$$0 \leq u(x) - \big(\langle \mathbf{p}_\alpha, x \rangle + a_\alpha\big) \leq \phi \left(\left(\frac{\eta}{\eta_{\phi,\widehat{\mathrm{R}}}} \right)^{\frac{1}{2k}} [|x| - \tau]_+ \right), \quad x \in \mathbf{B}_{\mathrm{R}}(x_0), \quad (4.22)$$

where

$$\tau = \left[\eta_*^{-\frac{1}{2k}} - \eta^{-\frac{1}{2k}} \right] \phi^{-1}(\mathrm{M}) \eta_{\phi,\widehat{\mathrm{R}}}^{\frac{1}{2k}}, \quad (4.23)$$

once we assume that $\mathrm{R} < \tau + \mathrm{R}_{\phi,\eta,\widehat{\mathrm{R}}}$ *and*

$$\left(\frac{\eta_{\phi,\widehat{\mathrm{R}}}}{\eta} \right)^{\frac{1}{2k}} \phi^{-1}(\mathrm{M}) < \mathrm{R} \leq \mathrm{dist}(x_0, \partial\Omega). \quad (4.24)$$

In particular, the function u is flat on $\overline{\mathbf{B}}_\tau(x_0)$. More precisely,

$$u(x) = \langle \mathbf{p}_\alpha, x\rangle + a_\alpha \quad \textit{for any } x \in \overline{\mathbf{B}}_\tau(x_0).$$

□

Remark 3 Corollary 1 is the relative version of Theorem 3. Consequently, the comments of Remarks 1 and 2 apply. □

In the particular case $f(t) = t^{\mathrm{q}}$, the condition (1.6) holds if and only if $k > \mathrm{q}$. Moreover, the assumption (4.7) is verified for

$$\mathbb{U}_{\mathrm{q},k}(r) = r^{\frac{2k}{k-\mathrm{q}}}, \quad \eta_{\mathrm{q},k} = \mathrm{C}_{\mathrm{N}-1,k-1}\left(\frac{2k}{k-\mathrm{q}}\right)^k \frac{2k\mathrm{q} + \mathrm{N}(k-\mathrm{q})}{k(k-\mathrm{q})}, \quad r \geq 0, \tag{4.25}$$

consequently all above results apply. If we scale by $\mathrm{A}^{\frac{k-\mathrm{q}}{2k}}$ for the function

$$\mathrm{U}(r) = \mathrm{A}\mathbb{U}_{\mathrm{q},k}(r), \quad r \geq 0,$$

the property (4.8) becomes

$$-\mathrm{C}_{\mathrm{N}-1,k-1} r^{1-\mathrm{N}}\left[\frac{r^{\mathrm{N}-k}}{k}\big(\mathrm{U}'(r)\big)^k\right]' + \eta \mathrm{U}(r)^{\mathrm{q}} = \eta\left[1 - \frac{\eta_{\mathrm{q},k}}{\eta}\mathrm{A}^{k-\mathrm{q}}\right]\mathrm{U}(r)^{\mathrm{q}} \tag{4.26}$$

fot $r > 0$. Now,

1. if $\mathrm{A} < \left(\frac{\eta}{\eta_{\mathrm{q},k}}\right)^{\frac{1}{k-\mathrm{q}}}$ the function $\mathrm{U}(r)$ is a super–solution of the equation (4.26),
2. if $\mathrm{A} = \left(\frac{\eta}{\eta_{\mathrm{q},k}}\right)^{\frac{1}{k-\mathrm{q}}}$ the function $\mathrm{U}(r)$ is the solution of the equation (4.26),
3. if $\mathrm{A} > \left(\frac{\eta}{\eta_{\mathrm{q},k}}\right)^{\frac{1}{k-\mathrm{q}}}$ the function $\mathrm{U}(r)$ is a sub–solution of the equation (4.26).

So that, the particular choice

$$\mathrm{U}(r) = \left(\frac{\eta}{\eta_{\mathrm{q},k}}\right)^{\frac{1}{k-\mathrm{q}}}\mathbb{U}_{\mathrm{q},k}(r), \quad r \geq 0, \tag{4.27}$$

enables us to construct the function

$$v_\tau(x) \doteq \mathrm{U}\big([|x| - \tau]_+\big), \quad x \in \mathbb{R}^{\mathrm{N}}, \tag{4.28}$$

vanishing in a ball $\mathbf{B}_\tau(0)$ and solving

$$-\mathcal{S}_k[\lambda(\mathrm{D}^2 v_\tau(x))] + \eta v_\tau(x)^{\mathrm{q}} = 0, \quad x \in \mathbb{R}^{\mathrm{N}}.$$

Moreover, given $\mathrm{M} > 0$, it verifies

$$v_\tau(x) = \mathrm{M}, \quad |x| = \mathrm{R}$$

once we take

$$\tau = \mathrm{R} - \mathrm{U}^{-1}(\mathrm{M}) = \eta_{\mathrm{q},k}^{\frac{1}{2k}} \mathrm{M}^{\frac{k-\mathrm{q}}{2k}} \left[\eta_*^{-\frac{1}{2k}} - \eta^{-\frac{1}{2k}} \right]$$

with

$$\eta \geq \eta_* \doteq \frac{\eta_{\mathrm{q},k} \mathrm{M}^{k-\mathrm{q}}}{\mathrm{R}^{2k}}.$$

The localization of a core of the flat region $\mathrm{Flat}(u)$ inside the flat subregion $\mathrm{Flat}_\alpha(h)$ of the "obstacle" is estimated by

Theorem 5 *Let $f(t) = t^{\mathrm{q}}$, $\mathrm{q} < k$. Let h be locally k–admissible on $\overline{\Omega}$. Let us assume that there exists $\mathbf{B}_{\mathrm{R}}(x_0) \subset \mathrm{Flat}_\alpha(h)$ with*

$$0 \leq u(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big) \leq \mathrm{M} \leq \max_{\overline{\Omega}}(u - h), \quad x \in \partial \mathbf{B}_{\mathrm{R}}(x_0), \tag{4.29}$$

where u is a generalized solution of (1.2), for some $\mathrm{M} > 0$. Then, if $k > \mathrm{q}$ and

$$\eta \geq \eta_* \doteq \frac{\eta_{\mathrm{q},k} \mathrm{M}^{k-\mathrm{q}}}{\mathrm{R}^{2k}}. \tag{4.30}$$

one verifies

$$0 \leq u(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big) \leq \left(\frac{\eta}{\eta_{\mathrm{q},k}}\right)^{\frac{1}{k-\mathrm{q}}} \big[|x - x_0| - \tau\big]_+^{\frac{2k}{k-\mathrm{q}}}, \quad x \in \mathbf{B}_{\mathrm{R}}(x_0), \tag{4.31}$$

where

$$\tau = \eta_{\mathrm{q},k}^{\frac{1}{2k}} \mathrm{M}^{\frac{k-\mathrm{q}}{2k}} \left[\eta_*^{-\frac{1}{2k}} - \eta^{-\frac{1}{2k}} \right] \tag{4.32}$$

provided

$$\left(\frac{\eta_{\mathrm{q},k}}{\eta}\right)^{\frac{1}{2k}} \mathrm{M}^{\frac{k-\mathrm{q}}{2k}} < \mathrm{R} \leq \mathrm{dist}(x_0, \partial\Omega). \tag{4.33}$$

In particular, the function u is flat on $\overline{\mathbf{B}}_\tau(x_0)$. More precisely,

$$u(x) = \langle \mathbf{p}_\alpha, x\rangle + a_\alpha \quad \textit{for any } x \in \overline{\mathbf{B}}_\tau(x_0).$$

□

Remark 4 Theorem 5 is a new version of Theorem 3 but now with more explicit data. Therefore, once more the comments of Remarks 1 and 2 apply also to this power like case $f(t) = t^{\mathrm{q}}$, $k > \mathrm{q}$. □

Theorem 5 gives some estimates on the localization of the points inside $\mathrm{Flat}(h)$ where u becomes flat too. The following result shows that if h decays in a suitable way at the boundary points of $\mathrm{Flat}(h)$ then the solution u becomes also flat in those points of the boundary of $\mathrm{Flat}(h)$. In this result the parameter η is irrelevant, therefore with no loss of generality we shall assume that $\eta = 1$.

Theorem 6 *Let* $f(t) = t^{\mathrm{q}}$, $\mathrm{q} < k$. *Let* $x_0 \in \partial\mathrm{Flat}_\alpha(h)$ *such that*

$$h(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big) \leq \mathrm{K}|x - x_0|^{\frac{2k}{k-\mathrm{q}}}, \quad x \in \mathbf{B}_{\mathrm{R}}(x_0) \cap \big(\mathbb{R}^{\mathrm{N}} \setminus \mathrm{Flat}(h)\big), \tag{4.34}$$

and

$$0 \leq \max_{|x-x_0|=\mathrm{R}} \big\{u(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big)\big\} \leq \mathrm{A}\mathrm{R}^{\frac{2k}{k-\mathrm{q}}} \tag{4.35}$$

for some suitable positive constants K *and* A (*see* (4.37) below) *and* u *is a generalized solution of* (1.2). *Then*

$$u(x_0) = \langle \mathbf{p}_\alpha, x_0\rangle + a_\alpha. \tag{4.36}$$

Proof. Define the function

$$\mathrm{V}(x) = u(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big),$$

which by construction is non–negative in $\partial\mathbf{B}_{\mathrm{R}}(x_0)$ (see (4.35)). In fact, the Weak Maximum Principle implies that V is non–negative on $\overline{\mathbf{B}}_{\mathrm{R}}(x_0)$. Then

$$\begin{aligned}
-\big(\mathcal{S}_k[\lambda(\mathrm{D}^2\mathrm{V}(x))]\big)^{\frac{1}{k}} + \mathrm{V}(x)^{\frac{\mathrm{q}}{k}} &= -\big(\mathcal{S}_k[\lambda(\mathrm{D}^2 u(x))]\big)^{\frac{1}{k}} + \big(u(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big)\big)^{\frac{\mathrm{q}}{k}} \\
&= -\big(u(x) - h(x)\big)^{\frac{\mathrm{q}}{k}} + \big(u(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big)\big)^{\frac{\mathrm{q}}{k}} \\
&\leq \big(h(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big)\big)^{\frac{\mathrm{q}}{k}} \\
&\leq \mathrm{K}^{\frac{\mathrm{q}}{k}}|x - x_0|^{\frac{2k}{k-\mathrm{q}}}, \quad x \in \mathbf{B}_{\mathrm{R}}(x_0),
\end{aligned}$$

where we have used a kind of Minkovsky inequality

$$(a+b)^{\frac{1}{\mathrm{p}}} \leq a^{\frac{1}{\mathrm{p}}} + b^{\frac{1}{\mathrm{p}}}, \text{ for any } a, b \geq 0, \text{ where } \mathrm{p} > 1,$$

for the special case $\mathrm{p} = \frac{k}{\mathrm{q}} > 1$, as well as (4.34). On the other hand, from (4.25) we have

$$\left(\mathrm{C}_{\mathrm{N}-1,k-1} r^{1-\mathrm{N}} \left[\frac{r^{\mathrm{N}-k}}{k}\big(\mathbb{U}'_{\mathrm{q},k}(r)\big)^k\right]'\right)^{\frac{1}{k}} = \eta_{\mathrm{q},k}^{\frac{1}{\mathrm{N}}} \mathbb{U}_{\mathrm{q},k}(r)^{\frac{\mathrm{q}}{\mathrm{N}}}, \quad 0 < r < \mathrm{R}_{\eta_{\mathrm{q},k}},$$

for

$$\mathbb{U}_{\mathrm{q},k}(r) = r^{\frac{2k}{k-\mathrm{q}}}, \quad \eta_{\mathrm{q},k} = \mathrm{C}_{\mathrm{N}-1,k-1}\left(\frac{2k}{k-\mathrm{q}}\right)^k \left(\frac{2k\mathrm{q} + \mathrm{N}(k-\mathrm{q})}{k(k-\mathrm{q})}\right), \quad \mathrm{R}_{\mathrm{q},k} = +\infty.$$

Then since the function $\mathrm{U}(r) = \mathrm{A}\mathbb{U}_{\mathrm{q},k}(r)$ verifies

$$-\mathrm{C}_{\mathrm{N}-1,k-1} r^{1-\mathrm{N}} \left[\frac{r^{\mathrm{N}-k}}{k} \big(\mathrm{U}'(r)\big)^k \right]' + \mathrm{U}(r)^{\mathrm{q}} = \left[1 - \eta_{\mathrm{q},k} \mathrm{A}^{k-\mathrm{q}}\right] \mathrm{U}(r)^{\mathrm{q}}, \quad 0 < r.$$

we may take $\mathrm{A} < \eta_{\mathrm{q},k}^{-\frac{1}{k-\mathrm{q}}}$ and then K such that

$$\mathrm{K}^{\frac{\mathrm{q}}{k}} \leq \mathrm{A}^{\frac{\mathrm{q}}{k}} \left[1 - \eta_{\mathrm{q},k} \mathrm{A}^{k-\mathrm{q}}\right]. \tag{4.37}$$

Then we obtain

$$-\big(\mathcal{S}_k[\lambda(\mathrm{D}^2\mathrm{V}(x))]\big)^{\frac{1}{k}} + \big(\mathrm{V}(x)\big)^{\frac{\mathrm{q}}{k}} \leq -\big(\mathcal{S}_k[\lambda(\mathrm{D}^2\mathrm{U}(|x|))]\big)^{\frac{1}{k}} + \mathrm{U}(|x|)^{\frac{\mathrm{q}}{k}}, \quad x \in \mathbf{B}_{\mathrm{R}}(x_0).$$

Finally, by choosing R satisfying (4.35) one has

$$\mathrm{V}(x) \leq \mathrm{U}(|x|), \quad x \in \partial \mathbf{B}_{\mathrm{R}}(x_0),$$

whence the comparison principle concludes

$$0 \leq \mathrm{V}(x) \leq \mathrm{A}|x - x_0|^{\frac{2k}{k-\mathrm{q}}}, \quad x \in \mathbf{B}_{\mathrm{R}}(x_0),$$

and so $u(x_0) = \big(\langle \mathbf{p}_\alpha, x_0\rangle + a_\alpha\big)$. □

Remark 5 The assumption (4.35) is satisfied if we know that the ball $\mathbf{B}_{\mathrm{R}}(x_0)$ where (4.34) holds is assumed large enough. The above result is motivated by [9, Theorem 2.5]. By adapting the reasoning used in previous results of the literature (see [8, Remark 10]) it can be shown that the decay of $h(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big)$ near the boundary point x_0 is optimal in the sense that if

$$h(x) - \big(\langle \mathbf{p}_\alpha, x\rangle + a_\alpha\big) > \mathrm{A}|x - x_0|^{\frac{2k}{k-\mathrm{q}}} \quad \text{on a neighbourhood of } x_0$$

then it can be shown that

$$u(x_0) - \big(\langle \mathbf{p}_\alpha, x_0\rangle + a_\alpha)\big) > \mathrm{A}|x - x_0|^{\frac{2k}{k-\mathrm{q}}} \quad \text{for } x \text{ near } x_0.$$

This type of results gives very rich information on the non–degeneracy behavior of the solution near the free boundary. This is very useful to the study of the continuous dependence of the free boundary with respect to the data h and φ (see [10]). □

5. Unflat solutions

The case where the free boundary can not appear (even if a priori the diffusion operator is degenerate) is examined here. Independent on the size of the domain, it requires the condition

$$k \leq \mathrm{q} \qquad \text{for } f(t) = t^{\mathrm{q}}$$

or the more general assumption (1.7). We shall obtain here a version of the Strong Maximum Principles inspired on the classical reasoning by E. Höpf (see *e.g.* [12, 16, 20]). Among other consequences, we shall deduce that the solution can not be flat. Again, since the parameter η is irrelevant, in this section, with no loss of generality, we assume $\eta = 1$. So, we begin with

Lemma 5.1 (Höpf boundary point lemma) *Assume* (1.7). *Let u be a non-negative viscosity solution of*

$$-\mathcal{S}_k[\lambda(\mathrm{D}^2 u)] + f(u) \geq 0 \quad \textit{in } \Omega.$$

Let $x_0 \in \partial\Omega$ be such that $u(x_0) \doteq \liminf\limits_{\substack{x \to x_0 \\ x \in \Omega}} u(x)$ and

$$\begin{cases} \textit{i)} & u \textit{ achieves a strict minimum on } \Omega \cup \{x_0\}, \\ \textit{ii)} & \exists\ \mathbf{B}_{\mathrm{R}}(x_0 - \mathrm{R}\mathbf{n}(x_0)) \subset \Omega \quad (\partial\Omega \textit{ satisfies an interior sphere condition at } x_0). \end{cases}$$

Then there exists a positive constant C such that

$$\liminf_{\tau \to 0} \frac{u(x_0 - \tau\mathbf{n})}{\tau} \geq C > 0, \tag{5.1}$$

where $\mathbf{n}$ stands for the outer normal unit vector of $\partial\Omega$ at x_0.

Proof. Let $y = x_0 - \mathrm{R}\mathbf{n}(x_0)$ and $\mathbf{B}_{\mathrm{R}} \doteq \mathbf{B}_{\mathrm{R}}(y)$. As it was pointed out before, the equation (1.2) leads to the study of the differential equation

$$\mathrm{C}_{\mathrm{N}-1,k-1}\left[\frac{r^{\mathrm{N}-k}}{k}\big(\Phi'(r)\big)^k\right]' = f\big(\Phi(r)\big), \quad r > 0,$$

for radially symmetric solutions. We consider now the classical solution of the boundary value problem

$$\begin{cases} \mathrm{C}_{\mathrm{N}-1,k-1} r^{1-\mathrm{N}}\left[\dfrac{r^{\mathrm{N}-k}}{k}\big(\Phi'(r)\big)^k\right]' = f\big(\Phi(r)\big), \quad 0 < r < \dfrac{\mathrm{R}}{2}, \\ \Phi(0) = 0, \quad \Phi\left(\dfrac{\mathrm{R}}{2}\right) = \Phi_1 > 0. \end{cases} \tag{5.2}$$

The existence of solution follows from standard arguments and the uniqueness of solution can be proved as in Theorem 2, whence

$$\Phi'(0) \geq 0 \quad \Rightarrow \quad \Phi'(r) > 0 \qquad \Rightarrow \quad \Phi''(r) > 0.$$

Then

$$0 \leq \Phi(r) \leq \Phi_1, \quad 0 < r < \frac{\mathrm{R}}{2}.$$

Obviously, the singularity at $r=0$ must be removed since we have

$$\lim_{r\to 0} r^{1-\mathrm{N}}\left[r^{\mathrm{N}-k}\big(\Phi'(r)\big)^k\right]' = 0. \tag{5.3}$$

Let r_0 be the largest r for which $\Phi(r)=0$. We want to prove that $r_0=0$ by proving that $r_0>0$ leads to a contradiction. In order to do that we note

$$\mathrm{C}_{\mathrm{N}-1,k-1}\Phi''(r)\left(\frac{\Phi'(r)}{r}\right)^{k-1} \le f\big(\Phi(r)\big), \quad 0<r<\frac{\mathrm{R}}{2}$$

(see (4.5)). So, we multiply by $r^{k-1}\Phi'(r)$ to get

$$\left[\big(\Phi'(r)\big)^{k+1}\right]' \le \mathrm{C}^{-1}_{\mathrm{N}-1,k-1}(k+1)f\big(\Phi(r)\big)\Phi'(r)r^{k-1}, \quad 0<r<\frac{\mathrm{R}}{2}.$$

Next, since $\Phi'(r_0)=0=\Phi(r_0)$, an integration between r_0 and r leads to

$$\begin{aligned}\big(\Phi'(r)\big)^{k+1} &\le \mathrm{C}^{-1}_{\mathrm{N}-1,k-1}(k+1)\mathrm{F}\big(\Phi(r)\big)r^{k-1}\\ &\quad -\mathrm{C}^{-1}_{\mathrm{N}-1,k-1}(k+1)(k-1)\int_{r_0}^{r}\mathrm{F}\big(\Phi(s)\big)r^{k-2}ds\\ &\le \mathrm{C}^{-1}_{\mathrm{N}-1,k-1}(k+1)\mathrm{F}\big(\Phi(r)\big)r^{k-1}, \quad r_0<r<\frac{\mathrm{R}}{2}.\end{aligned}$$

Since we assume (1.7), a new integration between r_0 and $\dfrac{\mathrm{R}}{2}$ yields

$$\begin{aligned}\infty=\int_0^{\Phi_1}\frac{ds}{\mathrm{F}(s)^{\frac{1}{k+1}}} &= \int_{r_0}^{\frac{\mathrm{R}}{2}}\frac{\Phi'(r)}{\mathrm{F}\big(\Phi(r)\big)^{\frac{1}{k+1}}}dr\\ &\le \Big(\mathrm{C}^{-1}_{\mathrm{N}-1,k-1}(k+1)\Big)^{\frac{1}{k+1}}\int_{r_0}^{\frac{\mathrm{R}}{2}} r^{\frac{k-1}{k+1}}dr<\infty\end{aligned}$$

and the conjectured contradiction follows. So that, we have proved $\Phi'(0)>0$ and also

$$0<\Phi(r)<\Phi_1,\ \Phi'(r)>0, \quad 0<r<\frac{\mathrm{R}}{2},$$

as well as $\Phi''(0)=0$ (see (5.3)). Hence, straightforward computations on the $\mathcal{C}^2$ convex function $w(x)=\Phi(\mathrm{R}-|x-y|)$, defined in the annulus $\mathcal{O}\doteq\mathbf{B}_{\mathrm{R}}\setminus\overline{\mathbf{B}}_{\frac{\mathrm{R}}{2}}$, prove

$$\begin{cases}\mathcal{S}_k[\lambda\big(\mathrm{D}^2w(x)\big)]=f\big(w(x)\big), & x\in\mathcal{O},\\ w(x)=\Phi_1, & x\in\partial\mathbf{B}_{\frac{\mathrm{R}}{2}},\\ w(x)=0, & x\in\partial\mathbf{B}_{\mathrm{R}}.\end{cases}$$

Moreover, by construction

$$u(x) > 0, \quad x \in \partial \mathbf{B}_{\frac{\mathrm{R}}{2}} \quad \Rightarrow \quad u(x) \geq w(x), \quad x \in \partial \mathbf{B}_{\mathrm{R}},$$

for Φ_1 small enough. Then the Weak Maximum Principle of Theorem 2 implies

$$(u - w)(x) \geq 0, \quad x \in \overline{\mathcal{O}}.$$

This leads to

$$\frac{u(x_0 - \tau \mathbf{n})}{\tau} \geq \frac{\Phi(\mathrm{R} - \mathrm{R}(1-\tau))}{\tau}, \qquad (\tau \ll 1)$$

whence

$$\liminf_{\tau \to 0} \frac{u(x_0 - \tau \mathbf{n})}{\tau} \geq \Phi'(0) > 0.$$

□

Remark 6 In fact, the above result implies

$$\liminf_{\substack{x \to x_0 \\ x \in \Omega}} \frac{u(x)}{|x - x_0|} \geq \Phi'(0) > 0.$$

□

If $\partial\Omega$ satisfies an uniform interior sphere condition one has

$$u(x) \geq \Phi'(0) \mathrm{dist}(x, \partial\Omega) \quad \text{near } \partial\Omega.$$

In particular, for $\Omega = \mathbf{B}_{\mathrm{R}}(y_0)$ we have

$$u(x) \geq \Phi'(0)(\mathrm{R} - |x - y_0|), \quad x \in \overline{\mathbf{B}}_{\mathrm{R}}(y_0). \tag{5.4}$$

This property can be improved

Theorem 7 (Locally lower bound) *Assume* (1.7). *Let u a positive viscosity solution u of*

$$-\mathcal{S}_k\big[\lambda\big(\mathrm{D}^2 u\big)\big] + f(u) \geq 0 \quad \text{in } \Omega.$$

Then, for each compact subset $\mathcal{K} \subset\subset \Omega$ there exists a positive constant $c_{\mathcal{K}}$ such that

$$u(x) \geq c_{\mathcal{K}}, \quad x \in \mathcal{K},$$

Proof. From the property (5.4) the conclusion holds for $\mathcal{K} \subset \mathbf{B}_{\mathrm{R}}(y_0) \subset\subset \Omega$. Next, the reasoning applies to every ball intersecting $\mathbf{B}_{\mathrm{R}}(y_0)$ and then to every ball that intersects one of those balls and so on. Finally, the conclusion holds for any compact $\mathcal{K} \subset\subset \Omega$, by means of suitable finite covering. □

Our main result proving the absence of the free boundary is the following

Theorem 8 (Höpf Strong Maximum Principle) *Assume* (1.7). *Let u be a non-negative viscosity solution of*

$$-\mathcal{S}_k[\lambda(\mathrm{D}^2 u)] + f(u) \geq 0 \quad \text{in } \Omega.$$

Then u can not vanish at some $x_0 \in \Omega$ unless u is constant in a neighborhood of x_0.

Proof. Assume that u is non–constant and achieves the minimum value $u(x_0)=0$ on some ball $\mathbf{B} \subset \Omega$. Then we consider the semi-concave approximation of u, *i.e.*

$$u^\varepsilon(x) \doteq \inf_{y\in\Omega} \left\{ u(y) + \frac{|x-y|^2}{2\varepsilon^2} \right\}, \quad x \in \mathbf{B}_\varepsilon \qquad (\varepsilon > 0), \tag{5.5}$$

where $\mathbf{B}_\varepsilon \doteq \{x \in \mathbf{B} : \ \mathrm{dist}(x, \partial\mathbf{B}) > \varepsilon\sqrt{1 + 4\sup_{\mathbf{B}} |u|}\}$. For ε small enough we can assume $x_0 \in \mathbf{B}_\varepsilon$. Then u^ε achieves the minimum value in $\mathbf{B}_\varepsilon$, with $u(x_0) = u^\varepsilon(x_0) = 0$. Moreover, from well known reasoning for general fully nonlinear equations (see, for instance [19, Proposition 2.3] or [1], [6]) one deduces that u^ε satisfies

$$-\mathcal{S}_k[\lambda(\mathrm{D}^2 u_\varepsilon)] + f\big(u_\varepsilon\big) \geq 0 \quad \text{on } \mathbf{B}_\varepsilon. \tag{5.6}$$

Denote

$$\mathbf{B}_\varepsilon^+ \doteq \{x \in \mathbf{B}_\varepsilon : \ u^\varepsilon(x) > 0\},$$

by geometric classical arguments there exists the largest ball $\mathbf{B}_{\mathrm{R}}(y) \subset \mathbf{B}_\varepsilon^+$ (see [12]). Certainly there exists some $z_0 \in \partial\mathbf{B}_{\mathrm{R}}(y) \cap \mathbf{B}_\varepsilon$ for which $u^\varepsilon(z_0) = 0$ is a local minimum. Then, Lemma 5.1 implies

$$\mathrm{D}u^\varepsilon(z_0) \neq \mathbf{0}$$

contrary to

$$\mathrm{D}u^\varepsilon(z_0) = \mathbf{0}, \tag{5.7}$$

(see Lemma 5.2 below). Therefore, u^ε is constant on $\mathbf{B} \subset \Omega$, *i.e.*

$$u^\varepsilon(y) = u^\varepsilon(x_0) = u(x_0), \quad y \in \mathbf{B}.$$

Finally, for every $y \in \mathbf{B}$ we denote by $\widehat{y}$ the point of Ω such that

$$u^\varepsilon(y) = u(\widehat{y}) + \frac{1}{2\varepsilon^2}|y - \widehat{y}|^2$$

whence

$$u(x_0) = u^\varepsilon(x_0) = u^\varepsilon(y) = u(y) + \frac{1}{2\varepsilon^2}|y-\widehat{y}|^2 \geq u(x_0) + \frac{1}{2\varepsilon^2}|y-\widehat{y}|^2 \geq u(x_0) \Rightarrow \widehat{y} = y.$$

So that, one concludes

$$u(y) = u^\varepsilon(y) = u^\varepsilon(x_0) = u(x_0), \quad y \in \mathbf{B}.$$

□

Corollary 2 *Assume* (1.7). *Let u be a generalized solution u of* (1.2). *Then if $u(x_0) > h(x_0)$ or $\mathcal{S}_k[\lambda(\mathrm{D}^2 h(x_0))] > 0$ at some point x_0 of a ball $\overline{\mathbf{B}} \subseteq \overline{\Omega}$ then $u > h$ on $\overline{\mathbf{B}}$, consequently the equation* (1.2) *is elliptic in $\overline{\mathbf{B}}$. In particular, if $\varphi(x_0) > h(x_0)$ at some $x_0 \in \partial\Omega$ or $\mathcal{S}_k[\lambda(\mathrm{D}^2 h(x_0))] > 0$ at some point $x_0 \in \Omega$ the problem* (3.6) *is elliptic non degenerate in path-connected open sets Ω, provided the compatibility condition* (1.4) *holds.*

Proof. From Theorem 8, both cases imply $u > h$ on $\overline{\mathbf{B}}$. Finally, a continuity argument concludes the proof. □

We end this section by proving property (5.7) used in the proof of Theorem 8

Lemma 5.2 *Let ψ be a function achieving a local minimum at some $z_0 \in \mathcal{O}$. Assume that there exists a function $\widehat{\psi}$ defined in $\mathcal{O}$ such that $\widehat{\psi}(z_0) = 0$, $\Psi = \psi + \widehat{\psi}$ is concave on $\mathcal{O}$ and*

$$\widehat{\psi}(x) \geq -\mathrm{K}|x - z_0|^2, \quad x \in \mathcal{O} \text{ with } |x - z_0| \text{ small,}$$

for some constant $\mathrm{K} > 0$. Then the function ψ is differentiable at z_0 and $\mathrm{D}\psi(z_0) = \mathbf{0}$.

Proof. By simplicity we can take $z_0 = 0 \in \mathcal{O}$. By applying the convex separation theorem there exists $\mathbf{p} \in \mathbb{R}^{\mathrm{N}}$ such that

$$\Psi(x) \leq \Psi(0) + \langle \mathbf{p}, x \rangle = \psi(0) + \langle \mathbf{p}, x \rangle, \quad x \in \mathcal{O}, \text{ with } |x| \text{ small.}$$

Then we have

$$\begin{aligned} \psi(x) &= \Psi(x) - \widehat{\psi}(x) \leq \psi(0) + \langle \mathbf{p}, x \rangle + \mathrm{K}|x|^2 \\ &\leq \psi(x) + \langle \mathbf{p}, x \rangle + \mathrm{K}|x|^2, \quad x \in \mathcal{O} \text{ with } |x| \text{ small} \end{aligned} \tag{5.8}$$

whence

$$-\langle \mathbf{p}, x \rangle \leq \mathrm{K}|x|^2, \quad x \in \mathcal{O} \text{ with } |x| \text{ small.}$$

For $\tau > 0$ small enough we can choose $x = -\tau\mathbf{p} \in \mathcal{O}$ and $\tau\mathrm{K} < 1$, for which

$$\tau|\mathbf{p}|^2 \leq \mathrm{K}\tau^2|\mathbf{p}|^2.$$

Therefore $\mathbf{p} = 0$. Finally, (5.8) leads to

$$0 \leq \psi(x) - \psi(0) \leq \mathrm{K}|x|^2, \quad x \in \mathcal{O} \text{ with } |x| \text{ small,}$$

and the result follows.

Remark 5.1 The result is immediate if ψ is concave (in this case we can choose $\widehat{\psi} \equiv 0$). The convex version follows by changing ψ and $\widehat{\psi}$ by $-\psi$ and $-\widehat{\psi}$, respectively (see Remark 3.2 above). □

Note that since the function u^ε defined in (5.5) is semi concave, the property (5.7) holds.

References

[1] G. Barles, J. Busca: *Existence and comparison results for fully nonlinear degenerate elliptic equations without zeroth-order term.* Comm. in P.D.E. **26** (11&12) (2001) 2323–2337.

[2] B. Brandolini, C. Trombetti: *Comparison results for Hessian equations via symmetrization.* J. Eur. Math. Soc. (JEMS) **9**(3) (2007) 561–575.

[3] L. Caffarelli, L. Nirenberg, J. Spruck: *The Dirichlet problem for nonlinear second-order elliptic equations III. Functions of eigenvalues of the Hessians.* Acta Math. **15** (1985) 261–301.

[4] K. S. Chou X.-J. Wang: *A variational theory of the Hessian equation.* Comm. Pure Appl. Math. **54** (2001) 1029–1064.

[5] A. Colesanti, P. Salani: *Generalized solutions of Hessian equations.* Bull. Austral. Math. Soc. **56** (1997) 459–466.

[6] M. G. Crandall, H. Ishii, P.-L. Lions: *User's guide to viscosity solutions of second order partial differential equations.* Bull. Amer. Math. Soc. **27** (1992) 1–67.

[7] G. Díaz: *Some properties of second order of degenerate second order P.D.E. in non-divergence form.* Appl. Anal. **20** (1985) 309–336.

[8] G. Díaz, J. I. Díaz: *On the free boundary associated with the stationary Monge-Ampère operator on the set of non strictly convex functions.* Discrete and Continuous Dynamical Systems **35**, 4 (2015) 1447–1468.

[9] J. I. Díaz: *Nonlinear Partial Differential Equations and Free Boundaries, Vol. 1, Elliptic Equations.* Res. Notes Math. **106,** Pitman, London 1985.

[10] J. I. Díaz, T. Mingazzini, A. M. Ramos: *On the optimal control for a semilinear equation with cost depending on the free boundary.* Networks and Heterogeneous Media **7**, 4 (2012) 605–615.

[11] D. Faraco, X. Zhong: *Quasiconvex functions and Hessian equations.* Arch. Rational Mech. Anal. **168** (2003) 245–252.

[12] D. Gilbarg, N. S. Trudinger: *Elliptic Partial Differential Equations of Second Order.* Springer-Verlag, Berlin 1983.

[13] M. E. Gurtin: *An Introduction to Continuum Mechanics.* Academic Press. New York 1981.

[14] C. E. Gutiérrez: *The Monge–Ampère equation.* Birkhauser, Boston, MA, 2001.

[15] R. Harvey, H. B. Lawson Jr.: *Calibrated geometries.* Acta Math. **148** (1982) 47–157.

[16] P. Pucci, J. Serrin: *The Maximum Principle.* Birkhäuser, Basel 2007.

[17] N. S. Trudinger: *The Dirichlet problem for the prescribed curvature equations.* Arch. Ration. Mech. Anal. **111** (1990) 153–179.

[18] N. S. Trudinger, X.-J. Wang: *Hessian measures.*I, Topol. Meth. Nonlinear Anal. **10** (1997) 225–239; II, Ann. Math. **150** (2) (1999) 579–604; III, J. Funct. Anal. **193** (1), 1–23.

[19] J. Urbas: *On the existence of nonclassical solutions for two classes of fully nonlinear elliptic equations.* Indiana Univ. Math. J. **39** (1990) 355–382.

[20] J. L. Vázquez: *A strong maximum principle for some quasilinear elliptic equations.* Appl. Math. Optim. **12** (1984) 191–202.

[21] X.-J. Wang: *The k-Hessian equation.* LNM **1997**, Springer-Verlag, Berlin 2009.

Hyperbolic surfaces of genus 3 with symmetry $\mathfrak{S}_4$

Raquel Díaz* **and Víctor González Aguilera**†

Departamento de Geometría y Topología, Fac. CC. Matemáticas
Universidad Complutense
28040 Madrid, Spain
radiaz@ucm.es

Departamento de Matemática
Universidad Técnica Federio Santa María
Valparaíso, Chile
victor.gonzalez@usm.cl

Dedicado a José María Montesinos en su 70 *cumpleaños.*

ABSTRACT

We consider the family $\mathcal{F}$ of hyperbolic surfaces of genus 3 with a group of isometries isomorphic to $\mathfrak{S}_4$. It is well known that this family is a one complex dimensional subvariety of $\mathcal{M}_3$ with three points in the boundary of $\mathcal{M}_3$.

In this note we find a subset $\mathcal{H}$ of Teichmüller space $\mathcal{T}_3$ covering $\mathcal{F}$. Using some specific Fenchel-Nielsen coordinates, this locus has a very simple description, and we find here some natural lines whose projections converge to the points in $\mathcal{F} \cap \partial\mathcal{M}_3$. We also give an explicit Fuchsian uniformization of the family of $\mathcal{H}$.

2010 Mathematics Subject Classification: 30F10, 14H15.
Key words: Hyperbolic surface, Teichmüller space, Fenchel-Nielsen coordinates.

*Author partially supported by project MTM2012-31973.
†Author supported by Fondecyt Project 1110516 and Dgyp of the UTFSM.

1. Introduction

A Riemann surface (of genus $g \geq 2$) can be studied from an algebraic point of view as a smooth complex algebraic curve, from an analytic point of view as a complex structure on a surface, or from a geometric point of view as a hyperbolic surface (i.e. a differential surface with a Riemannian metric of constant negative curvature). The set of all Riemann surfaces of genus g is called the *moduli space*, denoted $\mathcal{M}_g$. This space has a structure of analytic orbifold and its universal cover is the *Teichmüller space* of genus g, denoted $\mathcal{T}_g$.

Moduli space is not compact and it can be compactified adding *Riemann surfaces with nodes*, which are degenerations of Riemann surfaces when some system of curves collapse each one to a point called a *node* ([1], [7]). This compactification is denoted $\widehat{\mathcal{M}_g}$.

A Riemann surface with nodes can also be seen from the algebraic point of view as a stable curve [6]. A result of Deligne and Mumford states that $\widehat{\mathcal{M}_g}$ can be endowed with a structure of projective complex variety and contains $\mathcal{M}_g$ as a dense open subvariety [3].

In this paper we consider the family of Riemann surfaces of genus 3 admitting a subgroup of automorphisms isomorphic to the symmetric group $\mathfrak{S}_4$. From the algebraic point of view, this family has been studied in [8], and we will recall briefly their results at the end of this section. In this paper we are going to study the same family from the hyperbolic point of view. In Section 2 we will describe the locus $\mathcal{F}$ in $\mathcal{M}_3$ corresponding to this family and in Section 3 we will describe a locus $\mathcal{H}$ in $\mathcal{T}_3$ covering $\mathcal{F}$. Finally in Section 4 we will give a Fuchsian uniformization of the points in $\mathcal{H}$.

1.1. The KFT family

Let $\mathbb{P}^2(\mathbb{C})$ be the complex projective plane with homogeneous coordinates $(x_0 : x_1 : x_2)$ and let $\mathfrak{S}_4$ be the symmetric group of degree four. The homomorphism $\theta : \mathfrak{S}_4 \to \mathrm{Aut}(\mathbb{P}^2) \cong PGL(3, \mathbb{C})$ defined by $\theta((1234)) = A$, $\theta((12)) = B$, where

$$A(x_0 : x_1 : x_2) = (-x_1 : x_0 : -x_2) \qquad B(x_0 : x_1 : x_2) = (x_0 : x_2 : x_1)$$

determines a faithful representation.

The family of plane quartics invariant under the action of $\theta(\mathfrak{S}_4)$ is the pencil

$$KFT = \{X_{u,v} : u(x_0^4 + x_1^4 + x_2^4) + v(x_0^2x_1^2 + x_1^2x_2^2 +_3^2 x_0^2) = 0 \mid (u : v) \in \mathbb{P}^1(\mathbb{C})\}.$$

Each generic element of the pencil determines a non-hyperelliptic Riemann surface S (or a curve) of genus $g = 3$ such that $\theta(\mathfrak{S}_4) \subseteq \mathrm{Aut}(S)$.

This pencil contains the following four singular quartics [8]:

1. The curve $X_{(-2,1)}$ is the union of four lines:

$$\begin{array}{ll} l_1 : x_0 + x_1 - x_2 = 0; & l_2 : x_0 - x_1 + x_2 = 0 \\ l_3 : x_0 + x_1 + x_2 = 0; & l_4 : -x_0 + x_1 + x_2 = 0. \end{array}$$

We denote by $p_{ij} = l_i \cap l_j, i < j$. Thus, $X_{(-2,1)}$ is a Riemann surface with nodes the points p_{ij} and components l_i.

To each Riemann surface with nodes X we can associate a weighted graph $\mathcal{G}(X)$ whose vertices are the parts of X and there is an edge connecting to vertices if the parts corresponding to these vertices share one node. The weight of each vertex is the genus of the corresponding part.

In the case of $X_{(-2,1)}$, the associated graph $\mathcal{G}(X_{(-2,1)})$ is a tetrahedron.

2. The curve $X_{(-1,1)}$ is the union of two conics $C_1 \cup C_2$ where

$$C_1 : x_0^2 + \omega x_1^2 + \omega^2 x_2^2 = 0, \quad C_2 : x_0^2 + \omega^2 x_1^2 + \omega x_2^2 = 0$$

and ω is a primitive third root of the unity. The associated graph $\mathcal{G}(X_{(-1,1)})$ has two vertices C_1, C_2 and four edges, corresponding to the intersection points of C_1, C_2.

3. The curve $X_{(1,0)}$ is the irreducible quartic:

$$\{(x_0, x_1, x_2) \mid x_0^2 x_1^2 + x_1^2 x_2^2 + x_2^2 x_0^2 = 0\}$$

with three double points.

The associated graph is $\mathcal{G}(X_{(1,0)})$ has now one vertex and three loops corresponding to the three double points.

4. The quartic $X_{(2,1)}$ is a double conic which is a non reduced 1-dimensional subscheme of $\mathbb{P}^2(\mathbb{C})$.

It can be remarked that the quartic $X_{(2,1)}$ corresponds to the unique hyperelliptic Riemann surface in $\mathcal{M}_3$ with $\mathfrak{S}_4$ as reduced group of automorphisms. Therefore, the points of KFT in $\widehat{\mathcal{M}}_3 - \mathcal{M}_3$ are exactly $X_{(-2,1)}, X_{(-1,1)}, X_{(1,0)}$.

2. Hyperbolic surfaces invariant by $\mathfrak{S}_4$

In this section we construct hyperbolic surfaces $\Sigma_{l,t}$ with a subgroup of isometries isomorphic to $\mathfrak{S}_4$.

In the following, we will be using four oriented hyperbolic pairs of pants P_i, $i = 1, \ldots, 4$. The three boundary components of P_i are denoted by $\partial^i_j, \partial^i_k, \partial^i_m$, with $\{j, k, m\} = \{1, 2, 3, 4\} - \{i\}$. The common perpendicular between the component ∂^i_j and ∂^i_k is denoted by l^i_{jk}. See Figure 1.

We remark that, given two oriented hyperbolic pairs of pants P, P' each having a boundary component ∂, ∂' with common length l, in order to obtain an oriented

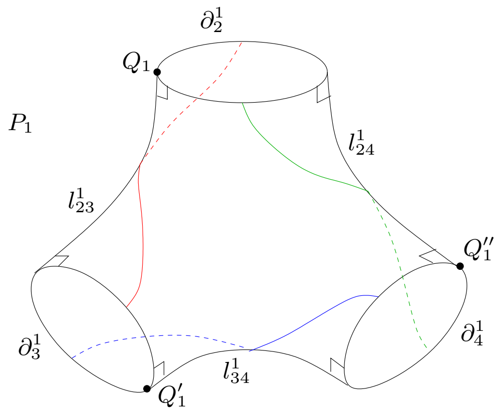


Figure 1: Pair of pants

hyperbolic surface by gluing ∂, ∂' we just need to indicate a single point in each of these boundary components to be glued together.

Consider the four hyperbolic pairs of pants $P_1, \ldots, P_4$ each of them having the lengths of its three boundary components equal to l. We identify each boundary component ∂_j^i with ∂_i^j so that the surface obtained be orientable and with the following data to determine the exact gluing:

a) (Initial case.) Glue P_1, P_2, P_3 and P_4 so that (i) the boundary perpendiculars $l_{23}^1, l_{13}^2, l_{21}^3$ match together to form a closed curve β_1; (ii) the boundary perpendiculars $l_{24}^1, l_{14}^2, l_{21}^4$ match together to form a closed curve β_2; and (iii) the boundary perpendiculars $l_{34}^1, l_{14}^3, l_{13}^4$ match together to form a closed curve β_3; automatically, the boundary perpendiculars $l_{34}^2, l_{24}^3, l_{23}^4$ also match together to form a closed curve β_4. We denote by $\Sigma_{l,0}$ the surface obtained. See Figure 2 (a).

b) (General case.) Starting from $\Sigma_{l,0}$, we perform a twist of width t in each gluing (see Figure 2(b)). We denote by $\Sigma_{l,t}$ the surface obtained. See Figure 2 (b).

We have the following theorem.

Theorem 2.1 *Let $\Sigma_{l,t}$ be the hyperbolic surface defined above.*

a) For each $l > 0, t \in \mathbb{R}$, the group $\mathfrak{S}_4$ acts on $\Sigma_{l,t}$ by hyperbolic isometries.

b) Conversely, let Σ be a hyperbolic surface admitting an action of $\mathfrak{S}_4$. Then Σ is isometric to $\Sigma_{l,t}$ for some $l > 0, t \in \mathbb{R}$.

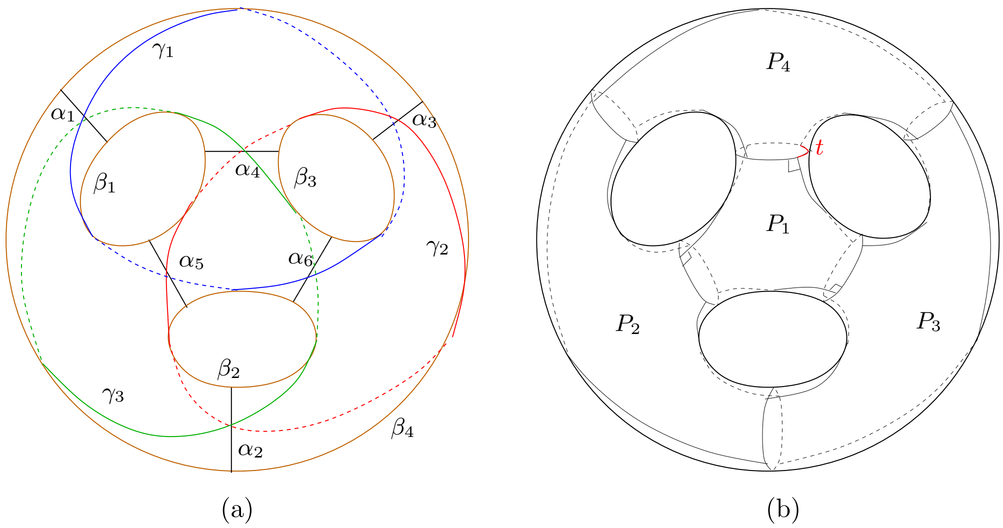


Figure 2: Gluing

Notation. We denote by $\mathcal{F}$ the family of hyperbolic surfaces invariant under $\mathfrak{S}_4$. Thus, the above theorem is proving that $\mathcal{F} = \{\Sigma_{l,t} : l > 0, 0 \leq t < l\}$.

Proof. a) Let σ be an element of $\mathfrak{S}_4$. For each pair of pants P_i we consider the orientation-preserving hyperbolic isometry $g_i^\sigma : P_i \to P_{\sigma(i)}$ so that $g_i(\partial^i_{jk}) = \partial^{\sigma(i)}_{\sigma(j)\sigma(k)}$. Notice that g_i^σ is well defined and uniquely determined by the previous conditions. The isometries g_i^σ are compatible with the identifications among the pants, giving rise to an isometry g^σ of $\Sigma_{l,t}$. This gives an action of $\mathfrak{S}_4$ on $\Sigma_{l,t}$ by isometries.

b) It is known ([2]) that there are exactly two non topologically equivalent actions of $\mathfrak{S}_4$ on a hyperbolic surface of genus 3 by hyperbolic isometries. In the first type of action there is a pants decomposition $\mathcal{A}$ invariant by the action, whose associated graph is isomorphic to the graph of the tetrahedron and the corresponding action on this graph is the usual action of $\mathfrak{S}_4$ in the edge graph of the tetrahedron. (A pants decomposition is a system of curves decomposing the surfaces in pants; to a pants decomposition we can associate a graph whose vertices are the pants and whose edges are the common boundary components). The quotient orbifold has signature $(0; 2, 2, 2, 3)$. The second type of action gives a quotient orbifold of signature $(0; 4, 4, 3)$. There is exactly one hyperbolic surface of genus 3 admitting the second action, this surface is hyperelliptic, and moreover it also admits another action of $\mathfrak{S}_4$ of the first type. In conclusion, any hyperbolic surface of genus 3 admitting an action of $\mathfrak{S}_4$, admits an action of $\mathfrak{S}_4$ of the first type.

Let Σ be a hyperbolic surface of genus 3, and let G be a subgroup of $Isom(\Sigma)$ isomorphic to $\mathfrak{S}_4$. Let $\mathcal{A}$ be a pants decomposition invariant under G whose associated graph is the edge graph of the tetrahedron. We denote by $P_1, \dots, P_4$ the pants in this decomposition and by $\partial^i_j (= \partial^j_i)$ the curve in $\mathcal{A}$ which is common boundary of P_i, P_j. Finally, we use the notation given in at the beginning of this section for the common perpendiculars to the boundaries of the pants.

Since G acts transitively on $\mathcal{A}$, then all these curves have the same length, say l. So, each pants has the three boundary components of length l.

Next we see how two of these pants are glued together. Assume first that P_1 is glued to P_2 so that the common perpendiculars l^1_{42} and l^2_{14} match together. We will see that this condition uniquely determines the other gluings. Let g be the element of G corresponding to the cycle $(124) \in \mathfrak{S}_4$; notice that g acts on the sets $\{P_i \colon i\}$, $\{\partial^i_j \colon i, j\}$ and $\{l^i_{jk} \colon i, j, k\}$ as indicated by the permutation (124) on the set of indices. In particular, l^1_{24} is mapped by g to l^2_{41}, and l^2_{14} is mapped to l^4_{21}. This shows that these three perpendiculars match up (to give a closed curve).

On the other hand, the fact that l^1_{42} and l^2_{14} match together implies that l^1_{23}, l^3_{12} do the same. And, since l^1_{42} and l^4_{13} match together, then also l^1_{43}, l^4_{13} match together. Then, applying similar arguments as before to the isometries corresponding to the cycles (123) and (134), we obtain that all the common perpendiculars match together to give four closed curves. This proves that the surface Σ is isometric to $\Sigma_{l,0}$.

In the general case, the pant P_1 is glued to P_2 so that the endpoints of the common perpendiculars l^1_{42} and l^2_{14} on ∂^1_2 are distant to each other by an amount of t to the left (as we look from the perpendicular to the boundary ∂^1_2); then, doing the same as before, this condition must be satisfied exactly in all the other gluings. Thus, the surface Σ is isometric to $\Sigma_{l,t}$. $\square$

3. Teichmüller space and Fenchel-Nielsen coordinates.

The construction in the previous section can be thought as finding some subset in the moduli space $\mathcal{M}_3$, the set of hyperbolic surfaces of genus 3 up to isometry.

If we want to compare the length of a fixed curve α as we move along different hyperbolic surfaces, we need to keep track of α along these different surfaces, i.e., we need to give a *marking* to each of them. If X is a hyperbolic surface of genus g, a marking on X is a homeomorphism $\phi \colon S_g \to X$, where S_g is a topological surface fixed from the beginning. The pair (X, ϕ) is called a *marked hyperbolic surface.* The *Teichmüller space* of genus g is the set $\mathcal{T}_g$ of equivalent classes of marked hyperbolic surfaces, where two such of them $(X_1, \phi_1), (X_2, \phi_2)$ are equivalent if there is an isometry $I \colon X_1 \to X_2$ so that $I \circ \phi_1$ is homotopic to ϕ_2. If α is a closed curve in S, the *length* of α in $\mathcal{X} = [(X, \phi)]$, denoted by $l_\alpha(\mathcal{X})$, is the length in X of the geodesic representing $\phi(\alpha)$. A multicurve $\mathcal{A} = \{\alpha_i\}$ is a family of simple disjoint closed curves, and the length of $\mathcal{A}$, denoted $l_{\mathcal{A}}$, means the sum of the lengths of the α_i.

An equivalent way to define Teichmüller space, that we will use later, is as the set of discrete, faithful representations $\rho \colon \pi_1(S) \to PSL(2, \mathbb{R})$ modulo conjugation by elements of $PGL(2, \mathbb{R})$.

The *mapping class group* of S is the group $MCG(S)$ of isotopy classes of homeomorphisms of S. An element f of $MCG(S)$ acts in $\mathcal{T}_g$ as follows:

$$f \cdot [(X, \phi)] = [(X, \phi \circ \bar{f}^{-1})],$$

where $\bar{f}$ is a representative of f. That is, the mapping class group acts on Teichmüller space by changing the markings. The quotient of $\mathcal{T}_g$ by this action is the moduli space $\mathcal{M}_g$.

The main example of an element of the mapping class group is a *Dehn twist* about a multicurve $\mathcal{A} = \{\alpha_i\}$, denoted $\tau_{\mathcal{A}}$. Intuitively, this homeomorphism consists on cutting along α_i, doing a whole twist of one of the two halves and gluing again.

Next we introduce Fenchel-Nielsen coordinates in $\mathcal{T}_3$ with respect to a reference $(\mathcal{A}, \mathcal{B})$. We will do it in the particular case we are interested on.

Let $\mathcal{A} = \{\alpha_1, \ldots, \alpha_6\}$ be a pants decomposition of S_3 as in the previous section, i.e., its associated graph is the edge graph of a tetrahedron. Let $\mathcal{B} = \{\beta_1, \ldots, \beta_4\}$ be the family of curves in S_3 drawn in Figure 2(a). Notice that these curves are obtained by gluing in a specific way arcs joining the boundaries of the four pairs of pants $S_3 - \mathcal{A}$.

We can parametrize $\mathcal{T}_g$ by the so called *Fenchel-Nielsen coordinates*, $(l_{\alpha_i}, t_{\alpha_i})$, where l_{α_i} are the lengths of the curves α_i and t_{α_i} are the *twist parameters*, which, intuitively, determine how the two pairs of pants meeting along the curve α_i are glued.

In order to define the twist parameters precisely, we need to specify, for each set of fixed values of the lengths of α_i, a base point $\mathcal{X}_0$ in Teichmüller space, at which all the twist parameters are equal to zero. Once we have this point, given a point $\mathcal{X} \in \mathcal{T}_g$ with l_{α_i} the given values, we can recover $\mathcal{X}_0$ from $\mathcal{X}$ by doing twists about α_i by certain amounts t_i. Then the twist parameters for $\mathcal{X}$ are defined to be $t_{\alpha_i} = -t_i$. To choose the base point, we will use the set of curves $\mathcal{B}$. For fixed values of l_{α_i}, the point $\mathcal{X}_0$ is chosen so that all the intersection angles between α_i and β_j are $\pi/2$.

After these preliminaries, we lift to $\mathcal{T}_3$ the family $\mathcal{F}$ of hyperbolic surfaces invariant by $\mathfrak{S}_4$ found in the previous section.

Corollary 3.1 *a) Let $\mathcal{H}$ be the subset of $\mathcal{T}_3$ with Fenchel-Nielsen coordinates $(l, \overset{(6}{\cdots}, l, t, \overset{(6}{\cdots}, t)$, $l > 0, t \in \mathbb{R}$, with respect to the marking $(\mathcal{A}, \mathcal{B})$. Then $\mathcal{H}$ projects onto $\mathcal{F}$.*

b) Let $\mathcal{A} = \{\alpha_1, \ldots, \alpha_6\}$, $\mathcal{B} = \{\beta_1, \ldots, \beta_4\}$ and $\mathcal{C} = \{\gamma_1, \gamma_2, \gamma_3\}$ be the curves in S drawn in Figure 2 (a). Then $\mathcal{A}, \mathcal{B}, \mathcal{C}$ are invariant under $\mathfrak{S}_4$.

c) Let $\mathcal{L}_1$ be the line in $\mathcal{H}$ defined by $t = 0$, and let $\mathcal{L}_2$ be the line in $\mathcal{H}$ defined by $t = l/2$. Then, $\mathcal{L}_1$ converges to the stable Riemann surface $X_{1,-2}$ when $l \to 0$, and to $X_{0,1}$ when $l \to \infty$. On the other hand, $\mathcal{L}_2$ converges to $X_{1,-1}$ when $t \to \infty$.

Remark. Notice that the multicurve $\mathcal{C}$ is a Borromean link embedded in S_3.

Proof. a) This is clear by the construction.

b) This is also clear. For the family $\mathcal{C}$, notice that these curves are made up by gluing the strings drawn in the pants of Figure 1.

c) If $t = 0$, $l \to 0$, then by definition the length of α_i tends to zero so in moduli space this converges to a hyperbolic surface with six ideal points and with the configuration of the tetrahedron. If $t = 0, l \to \infty$, then by the trigonometric formulae for a right angle hexagon, the common perpendiculars to the boundary of the pants tend to zero. Thus, the length of the curves β_i tend to zero, and the projection in moduli space converges to $X_{0,1}$. If $t = l/2$, then notice that the common perpendiculars to the boundaries of the pants match together but in a different way: l_{34}^1 matches with l_{12}^4, etc. After doing all the gluings, these arcs form the curves $\gamma_1, \gamma_2, \gamma_3$. Therefore, when $l \to \infty$, the length of the common perpendiculars tend to zero, and so do the lengths of γ_i. Thus, in moduli space it projects to $X_{1,-1}$. □

3.1. Action of the mapping class group.

The full preimage of $\mathcal{F}$ on $\mathcal{T}_3$ is the orbit of $\mathcal{H}$ by the mapping class group, which is the set of points in $\mathcal{T}_3$ fixed by the action of some subgroup of $MCG(S)$ isomorphic to $\mathfrak{S}_4$. The set $\mathcal{H}$ can be characterized by the set of points fixed by the subgroup G_0 of $MCG(S)$ isomorphic to $\mathfrak{S}_4$ and which leaves $\mathcal{A}$ invariant.

Theorem 3.1 *Let $\mathcal{A} = \{\alpha_1, ..., \alpha_6\}$, $\mathcal{B} = \{\beta_1, ..., \beta_4\}$ and $\mathcal{C} = \{\gamma_1, \gamma_2, \gamma_3\}$ the multicurves drawn in Figure 2(a). Let $\tau_{\mathcal{A}}, \tau_{\mathcal{B}}, \tau_{\mathcal{C}}$ be Dehn twist about respectively the multicurve $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$. Then $\tau_{\mathcal{A}}, \tau_{\mathcal{B}}, \tau_{\mathcal{C}}$ leave $\mathcal{H}$ invariant.*

Proof. Let us see it for $\tau_{\mathcal{C}}$. Let $\mathcal{X}$ be a point in $\mathcal{H}$ and let $(l_1, \ldots, l_6, t_1, \ldots, t_6)$ be the Fenchel-Nielsen coordinates with respect to $(\mathcal{A}, \mathcal{B})$ of the point $\tau_{\mathcal{C}} \cdot \mathcal{X}$. We need to prove that $l_1 = \ldots = l_6$ and $t_1 = \ldots = t_6$. We have

$$l_i = l_{\alpha_i}(\tau_{\mathcal{C}} \cdot \mathcal{X}) = l_{\tau_{\mathcal{C}}^{-1}(\alpha_i)}(\mathcal{X}).$$

Let us denote $\alpha_i' = \tau_{\mathcal{C}}^{-1}(\alpha_i)$. Since $\mathcal{A}$ and $\mathcal{C}$ are invariant by G_0, then $\mathcal{A}' = \{\alpha_1', \ldots, \alpha_6'\}$ is also invariant by G_0. Thus, $l_1 = \ldots = l_6$. A similar argument shows that $t_1 = \ldots = t_6$, since the twist coordinates are measured as distances on α_i of certain common perpendiculars; since the set of these elements are invariant by G_0, the result holds. □

Thus, the stabilizer of $\mathcal{H}$ in $MCG(S)$ contains the subgroup generated by $\tau_{\mathcal{A}}, \tau_{\mathcal{B}}, \tau_{\mathcal{C}}$.

Finally, we remark that it can be shown that $\mathcal{L}_1$ is the line of minima determined by $\mathcal{A}$, $\mathcal{B}$, and the $\mathcal{L}_2$ is the line of minima determined by $\mathcal{A}$, $\mathcal{C}$. Lines of minima were introduced by Kerckhoff [5] as curves in $\mathcal{T}_g$ analogous to geodesics in hyperbolic space. Under certain conditions on two multicurves $\mathcal{A}, \mathcal{B}$, the function $\varphi_s = sl_{\mathcal{A}} + (1-s)l_{\mathcal{B}}$ has a unique minimum. The line of minima determined by $\mathcal{A}, \mathcal{B}$ is the collection of such minima, as $0 < s < 1$ (the definition is more general, but this is enough for our purposes).

4. Hyperbolic representation of $\mathcal{H}$.

We provide an explicit representation for the hyperbolic surfaces in our family in terms of the Fenchel-Nielsen coordinates with respect to reference $(\mathcal{A}, \mathcal{B})$.
Notation. If R is an oriented hyperbolic line in $\mathbb{H}^2$ and a is a real number, we denote by $\bar{T}_R(a)$ the hyperbolic translation along R with translation distance equal to a (following the orientation of R when $a > 0$).

We take as generators of $\pi_1(S_3)$ the curves $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3$ shown in the left-hand side of Figure 4. Cutting $\Sigma_{l,t}$ along the pants curves $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2$ and developing on the hyperbolic plane, we obtain the right-hand side of Figure 4. We arrange this figure so that the line L is the hyperbolic geodesic going from 0 to ∞ (in the Poincaré half-plane model) and M is the geodesic with endpoints from -1 to 1. The hyperbolic translations along L, M thus have simple form, they are represented, respectively, by the matrices of $SL(2,\mathbb{R})$

$$T_L(a) = \begin{pmatrix} e^{a/2} & 0 \\ 0 & e^{-a/2} \end{pmatrix} \qquad T_M(a) = \begin{pmatrix} \cosh \frac{a}{2} & \sinh \frac{a}{2} \\ \sinh \frac{a}{2} & \cosh \frac{a}{2} \end{pmatrix}$$

We will express $\rho(\alpha_i)$ and $\rho(\beta_i)$ as products of hyperbolic translations in L and M.

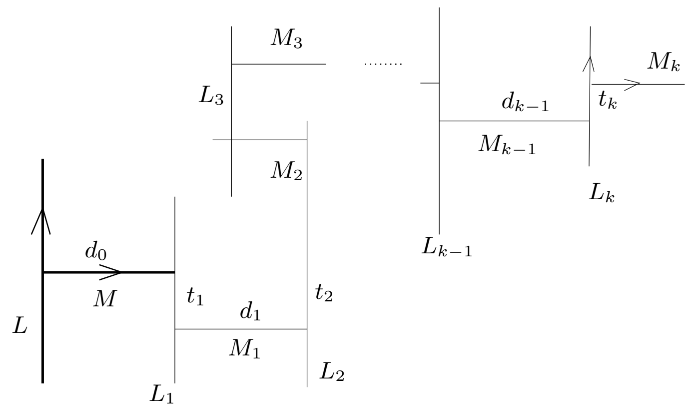


Figure 3: Broken arc

In order to do this, the following lemma will be very convenient.

Lemma 4.1 *Let $L = L_0, M = M_0, L_1, M_1, \ldots, L_k, M_k$ be hyperbolic lines so that each one is perpendicular to the previous and posterior ones and L, M are oriented.*

We schematically draw them so that the L_i's are vertical and the M_i's are horizontal and we orientate the L_i's equal as L and the M_i's equal as M. Let d_i be the signed distance between L_i and L_{i+1}, where the sign is positive if the hyperbolic translation along M_i taking L_i to L_{i+1} follows the orientation of M_i. Similarly, let t_i be the signed distance between M_{i-1} and M_i (see Figure 3). Let φ be the hyperbolic isometry taking the oriented pair (L, M) to the oriented pair (L_k, M_k). Then

$$\begin{aligned} \varphi &= \bar{T}_{L_k}(t_k) \circ \ldots \circ \bar{T}_{M_1}(d_1)\bar{T}_{L_1}(t_1) \circ \bar{T}_{M_0}(d_0) \\ &= \bar{T}_M(d_0) \circ \bar{T}_L(t_1) \circ \bar{T}_M(d_1) \circ \ldots \circ \bar{T}_M(d_{k-1}) \circ \bar{T}_L(t_k). \end{aligned}$$

Proof. We denote by P the polygonal line in the statement of the lemma. Let $\Phi = \bar{T}_M(d_0) \circ \bar{T}_L(t_1) \circ \bar{T}_M(d_1) \circ \ldots \circ \bar{T}_M(d_{k-1}) \circ \bar{T}_L(t_k)$. We need to show that $\Phi(L) = L_k$, $\Phi(M) = M_k$, which is equivalent to prove that $\Phi^{-1}(L_k) = L$, $\Phi^{-1}(M_k) = M$. Now,

$$\Phi^{-1} = \bar{T}_L(-t_k) \circ \bar{T}_M(-d_{k-1}) \ldots \bar{T}_M(-d_1) \circ \bar{T}_L(-t_1) \circ \bar{T}_M(-d_0).$$

We look at how Φ^{-1} is moving the polygonal $P = M_0, \ldots, M_k$: $\bar{T}_M(-d_0)$ moves P so that L_1 is mapped to L; $\bar{T}_L(-t_1) \circ \bar{T}_M(-d_0)$ moves P so that (M_1, L_1) is mapped to (M, L). Proceeding recursively in this way, we obtain the result. $\square$

Looking at Figure 4, we have:

$$\begin{aligned} &\rho(\alpha_1) = G_1\bar{T}_L(-l)G_1^{-1} & \rho(\beta_1) &= G_4(\bar{T}_M(d)\bar{T}_L(t))^3G_4^{-1} \\ &\rho(\alpha_2) = G_2\bar{T}_L(l)G_2^{-1} & \rho(\beta_2)^{-1} &= G_5(\bar{T}_M(d)\bar{T}_L(t))^3G_5^{-1} = (\bar{T}_M(d)\bar{T}_L(t))^3 \\ &\rho(\alpha_3) = G_3\bar{T}_L(-l)G_3^{-1} & \rho(\beta_3) &= G_6(\bar{T}_M(-d)\bar{T}_L(-t))^3G_6^{-1} \end{aligned}$$

where d is the length of the remaining edges in a right angle hyperbolic hexagon with three non-consecutive edges equal to l; and $G_1, \ldots, G_6$ are hyperbolic isometries so that

(i) G_1 maps the line L to the line L_5, through the polygonal line L, M_4, L_4, M_5, L_5

(ii) G_2 maps L to L_1 through the polygonal line L, M_1, L_1

(iii) G_3 maps L to L_7 through the polygonal line L, M, L_2, M_9, L_7

(iv) G_4 maps the oriented lines (L, M) to the oriented lines (L_0, M_0) ;

(v) G_5 maps (L, M) to (L_1, M_1);

(vi) G_6 maps (L, M) to (L_6, M_7);

The precise expressions for the G_i can be obtained using Lemma 4.1.

Remarks. (a) The above process can be done to obtain a representation for any point in $\mathcal{T}_3$. It is enough to change in Figure 4 the parameters $(l, \overset{(6}{\cdots}, l, t, \overset{(6}{\cdots}, t)$ by

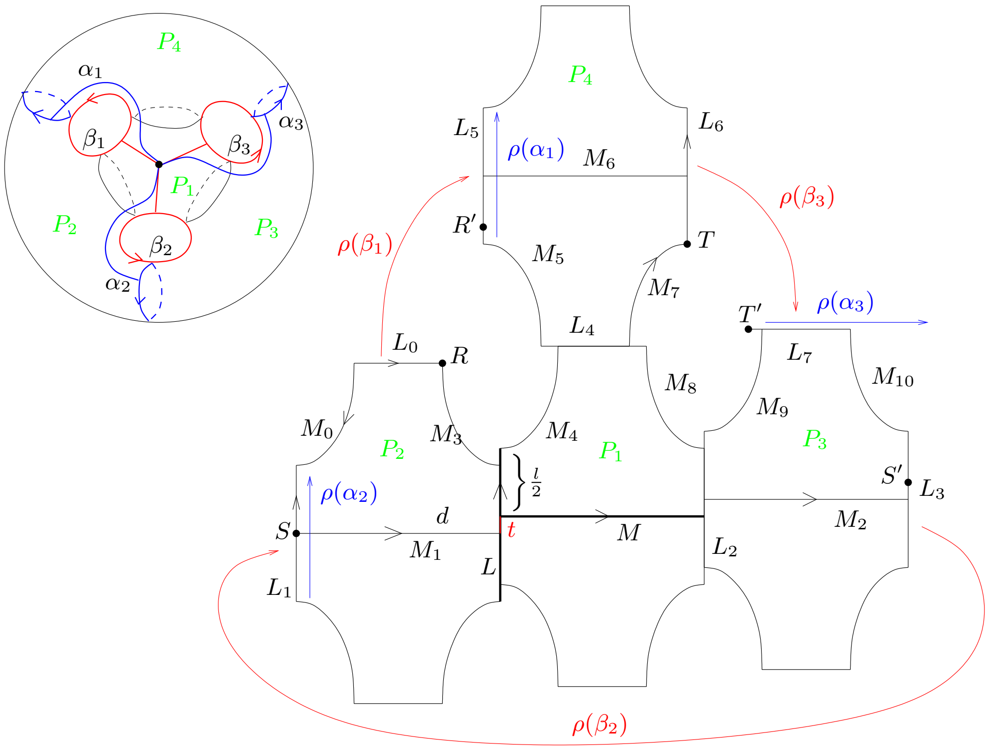


Figure 4: Presentation of the fundamental group and representation

$(l_1, \dots, l_6, t_1, \dots, t_6)$ and accordingly change the length d of the common perpendiculars by $d_1, \dots, d_9$. This type of representation was used in [4] to obtain the representation of quasi-Fuchsian space of the twice-punctured torus, and is convenient for computations.

(b) Another Fuchsian representation of the same family is obtained in [8] starting from a fundamental polygon of the quotient orbifold $(0; 2, 2, 2, 3)$.

References

[1] L. Bers: *On spaces of Riemann surfaces with nodes.* Bull. AMS **80** (1974) 1219–1222.

[2] S. A. Broughton: *Classifying finite group actions on surfaces of low genus.* J. Pure Appl. Algebra **69** (1990) 233–270.

[3] P. Deligne, D. Mumford: *The irreducibility of the space of curves.* Publ. Math. IHES **36** (1965) 75–109.

[4] R. Díaz, C. Series: *Examples of pleating varieties for twice punctured tori.* Trans. AMS **356**(2) (2003) 621–658.

[5] S. Kerckhoff: *Lines of Minima in Teichmüller space.* Duke Math. J. **65** (1992) 187–213.

[6] R. Miranda: *Graph curves and curves on K3 surfaces.* International Centre for Theoretical Physics, Trieste, World Scientific (1989) 119–176.

[7] Y. Matsumoto, J. M. Montesinos-Amilibia: *Pseudo-periodic maps and degenerations of Riemann surfaces.* LNM **2030** Springer-Verlag 2011.

[8] R. E. Rodríguez, V. González-Aguilera: *Fermat's Quartic curve, Klein's curve and the Tetrahedron.* Contemporary Mathematics **201** (1997) 43–61.

Modelo plano conforme del Plano Proyectivo

Fernando ETAYO y Ujué R. TRÍAS

Departmento de Matemáticas, Estadística y Computación
Universidad de Cantabria
39005 Santander, Spain
etayof@unican.es ujuert@gmail.com

Al profesor Montesinos, de cuyo magisterio nos hemos gozado,
y cuyo traslado a Madrid propició que dos líneas de vida concurrieran.

ABSTRACT

En esta nota presentamos un modelo plano conforme del Plano Proyectivo, similar al del disco de Poincaré del Plano Hiperbólico. Ambos modelos han sido usados [3] para definir herramientas proyectivas e hiperbólicas en el programa GeoGebra.

In this note, we present a conformal model of the Projective plane, similar to the Poincaré disk model of the Hyperbolic Plane. This model (and that of the Hyperbolic plane) have been used [3] to achieve hyperbolic and projective tools in the visualization program GeoGebra.

2010 Mathematics Subject Classification: 51M09, 51M10.
Key words: Projective plane, hyperbolic plane, models, Poincaré Disk.

1. Introducción

En una memoria de 1887, el matemático francés Henri Poincaré (1854-1912) describió un modelo concreto de una Geometría Hiperbólica en dos dimensiones; este modelo es conocido ahora como el disco de Poincaré. Los puntos en el modelo de Poincaré del Plano Hiperbólico son los puntos del interior de un círculo, y las rectas son aquellos arcos de circunferencia generalizada que intersecan ortogonalmente con la frontera del círculo, la circunferencia H. Se puede dotar a este modelo de Plano Hiperbólico con una medida de longitud no acotada, de tal manera que las rectas anteriormente definidas son las geodésicas del espacio métrico así generado. Los ángulos son medidos por sus valores como ángulos euclídeos.

En cuanto a la Geometría Esférica o Proyectiva, el modelo más natural de Plano Proyectivo lo constituye la propia esfera, tomando como puntos las parejas de puntos antipodales y como rectas las circunferencia maximales, que son las geodésicas de la esfera vista como superficie inmersa en el espacio Euclídeo tridimensional. La distancia entre dos puntos vendría dada por la longitud del arco de geodésica más corto que une ambos puntos. Nótese que esta métrica está acotada y tiene como valor máximo $\pi/2$.

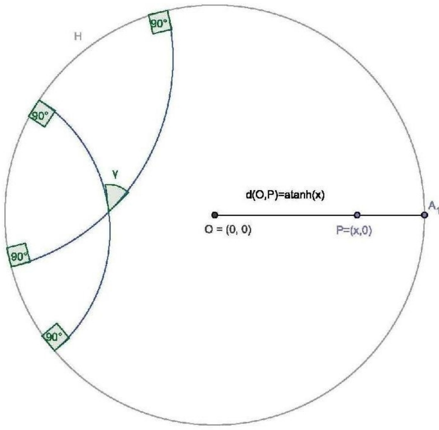
Nuestra idea fundamental es presentar un modelo del Plano Proyectivo similar en sus propiedades al del disco de Poincaré del Plano Hiperbólico que hemos mencionado. Esta nota es una versión abreviada de parte del trabajo [3], en el que además habíamos realizado todas las construcciones que permiten implementar con el programa GeoGebra ambos modelos. Respecto de las referencias sobre el modelo del Plano Hiperbólico son muchas las obras que se pueden consultar. Recomendamos [1] por su riqueza geométrica. Entendemos que la geometría del disco de Poincaré es conocida (en el propio trabajo [3] se contiene información detallada). El énfasis del trabajo no está en señalar las propiedades del plano proyectivo, sino específicamente las del modelo que vamos a dar del mismo.

2. Modelo plano y conforme de la Geometría Proyectiva

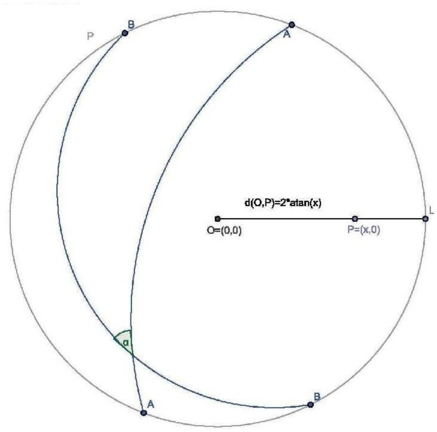
Partamos, pues, del Plano Proyectivo como cociente de la esfera, mediante la identificación de los puntos antipodales. Podemos construir un modelo plano y conforme mediante la proyección estereográfica desde el polo sur sobre el plano ecuatorial. Como todo punto proyectivo tiene al menos un representante en el hemisferio superior (incluido el ecuador), el disco unidad cerrado contiene en su interior un representante de cada punto no ecuatorial y su frontera, la circunferencia unidad P, contiene a los dos representantes de cada punto del ecuador.

Así obtenemos un modelo del Plano Proyectivo cuyos puntos son los puntos del interior del disco unidad más las parejas de puntos antipodales de la circunferencia unidad; sus rectas vienen dadas por los arcos de circunferencia generalizada que cortan a la circunferencia unidad en puntos diametralmente opuestos más la propia circunferencia unidad. Nótese que todas lar rectas proyectivas son líneas cerradas de longitud π.

Nota: La proyección estereográfica de $S^2 = \{(x, y, z)|x^2 + y^2 + z^2 = 1\}$, la esfera de radio 1, desde el polo sur $S = (0, 0, -1)$ sobre el plano $\{z = 0\}$ coincide con la restricción a S^2 de la inversión del espacio con respecto a la esfera de ecuación $\{x^2 + y^2 + (z + 1)^2 = 2\}$. En efecto, dicha inversión transforma esferas generalizadas en esferas generalizadas, envía el polo sur $S = (0, 0, -1)$ al infinito, con lo que S^2 se transforma en un plano y deja invariante la intersección $S^2 \cap \{x^2 + y^2 + (z + 1)^2 = 2\}$, esto es, la circunferencia unidad en el plano $\{z = 0\}$. Esto nos permite



DISCO DE POINCARÉ



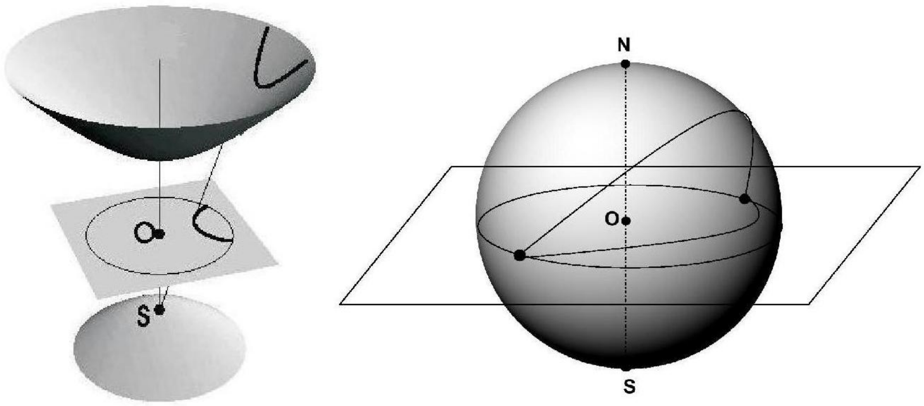
DISCO PROYECTIVO

afirmar que es una aplicación conforme que transforma circunferencias generalizadas en circunferencias generalizadas.

3. Punto de vista unificador

Es sabido que si consideramos el espacio tridimensional real dotado con la métrica de Lorentz en donde la distancia entre dos puntos $A = (a_1, a_2, a_3)$ y $B = (b_1, b_2, b_3)$ viene dada por $d(A, B)^2 = (b_1 - a_1)^2 + (b_2 - a_2)^2 - (b_3 - a_3)^2$, la esfera de radio $i = \sqrt{-1}$ centrada en el origen $O = (0, 0, 0)$, esto es, el conjunto de puntos del espacio tridimensional a distancia $i = \sqrt{-1}$ del origen O, es un hiperboloide equilátero de dos hojas que hereda una métrica Riemaniana de curvatura constante $K = -1$ y por tanto cada una de sus componentes conexas constituye un modelo de Espacio Hiperbólico bidimensional. Además, como en este caso el hiperboloide tiene intersección vacía con el plano $\{z = 0\}$, la identificación de puntos antipodales no añade ni quita nada al modelo que supone cada una de sus hojas por separado.

Pues bien, si ahora consideramos la proyección estereográfica desde el polo sur $S = (0, 0, -1)$ sobre el plano $\{z = 0\}$ de la hoja superior del hiperboloide, recuperamos el modelo del disco de Poincaré del Plano Hiperbólico. Las rectas de la hoja superior del hiperboloide, curvas generadas al intersecar este último con planos que pasan por el origen $O = (0, 0, 0)$, se proyectan en arcos de circunferencia del interior del disco unidad ortogonales a la circunferencia frontera de dicho disco. Es más, esta proyección estereográfica restringida al hiperboloide equilátero unitario de dos hojas puede verse como una inversión del espacio de Lorentz con respecto a la esfera de centro $S = (0, 0, -1)$ y de radio $\sqrt{-2}$, esto es, como la restricción al hiperboloide de la inversión del espacio con respecto al hiperboloide equilátero de ecuación $\{x^2 + y^2 - (z+1)^2 = -2\}$, si medimos las distancias con la métrica de Lorentz. Obtenemos así una clara analogía con el modelo del disco del Plano Proyectivo.



4. Sobre el Plano Proyectivo y su representación en el disco unidad

Es este apartado veremos algunas peculiaridades del Plano Proyectivo y cómo se plasman en el modelo del disco unidad. Los detalles pueden consultarse en [3]. Por ser el Plano Proyectivo un espacio cociente obtenido por la identificación de puntos antipodales de la esfera unidad, comenzaremos viendo algunas peculiaridades de la esfera como espacio topológico métrico y cómo pasan éstas al Plano Proyectivo.

- **Función distancia**. En la esfera la función distancia está acotada por el valor π que se alcanza sólo en el caso de puntos antipodales. Al identificar puntos antipodales para obtener como espacio cociente el Plano Proyectivo, debemos definir la función distancia a partir de la función de distancia en la esfera. Así, si A y B son dos puntos del Plano Proyectivo y A_1 y A_2, B_1 y B_2 son los puntos de la esfera cuyas clases de equivalencia mediante la identificación antipodal son A y B respectivamente, entonces definimos la distancia entre A y B como sigue:

$$d(A, B) = min[\delta(A_i, B_j)]i, j \in \{1, 2\}$$

en donde δ denota la distancia esférica. Esta función distancia también está acotada pero en este caso el valor máximo que alcanza es $\pi/2$.

- **Circunferencias**. Cabe preguntarse ahora cuál es el lugar geométrico de los puntos del Plano Proyectivo situados a distancia proyectiva $\pi/2$ de un punto dado A. De la propia definición deducimos que son los puntos del Plano Proyectivo que provienen de los puntos de la esfera cuya distancia δ a los puntos A_1 y A_2, cuya clase de equivalencia mediante la identificación antipodal es A, es $\pi/2$; esto es, los puntos de la circunferencia maximal contenida en el plano perpendicular al diámetro de extremos A_1 y A_2 que pasa por el centro de la esfera.

  Llamaremos a esta circunferencia maximal *circunferencia polar* de los puntos A_1 y A_2, y llamaremos *recta polar* de A a la recta proyectiva a que da lugar. Podemos responder pues, que el lugar geométrico de los puntos del Plano Proyectivo situados a distancia $\pi/2$ de un punto dado A es su recta polar. Hagamos notar, llegados a este punto, que como toda circunferencia maximal de la esfera unidad que pasa por los puntos antipodales A_1 y A_2 corta ortogonalmente a su circunferencia polar, toda recta proyectiva que pase por A cortará ortogonalmente a su recta polar.

- **Polo de una recta**. Recíprocamente, dada una recta proyectiva, el lugar geométrico de los puntos del Plano Proyectivo situados a distancia proyectiva $\pi/2$ de todos los puntos de la recta dada es un único punto, el punto de intersección de todas las polares de los puntos de la recta proyectiva dada. Llamaremos a este punto *polo* de la recta en cuestión. De hecho, la aplicación que a cada punto del Plano Proyectivo le hace corresponder su recta polar es una *correlación*.

- **Distancia entre rectas**. La correlación definida en el punto anterior nos permite definir una función distancia en el conjunto de rectas del Plano Proyectivo. En efecto si a y b son dos rectas proyectivas y A y B son sus polos respectivos definimos la distancia proyectiva entre a y b como la distancia proyectiva entre A y B. ¿Qué significado geométrico tiene esta función distancia? La distancia proyectiva entre dos rectas del Plano Proyectivo es igual a la medida en radianes del menor de los ángulos que forman.

- **El Plano Proyectivo como espacio homogéneo**. El grupo de isometrías del Plano Proyectivo actúa sobre el mismo transitivamente, esto es, dados dos puntos cualesquiera A y B del Plano Proyectivo, existe al menos una isometría del mismo que transforma A en B. Así el Plano Proyectivo es un espacio homogéneo. Aunque los puntos del Plano Proyectivo considerado en abstracto son indistinguibles por ser éste un espacio homogéneo, en el modelo del disco unidad con puntos antipodales de la circunferencia frontera identificados existen puntos distinguidos; a saber, el centro $O = (0, 0, 0)$, que tiene la propiedad de que las rectas que pasan por él se representan por diámetros del disco unidad con extremos identificados, y los puntos de la circunferencia frontera del disco por aparecer duplicados en el modelo. La singularidad del centro del disco es de gran utilidad a la hora de construir herramientas en el Plano Proyectivo pues nos permite construirlas en el centro y luego trasladarlas mediante una isometría a cualquier otro punto del Plano Proyectivo.

- **Isometrías del Plano Proyectivo**. El grupo de isometrías del Plano Proyectivo es el grupo cociente del grupo de isometrías de la esfera entre el subgrupo generado por la reflexión en el centro de la esfera, que únicamente contiene dos elementos: la reflexión antipodal y la identidad. El grupo de isometrías de la esfera está generado por las *reflexiones en circunferencias maximales*, esto es, por la restricción a la esfera de reflexiones en planos que pasan por el centro de la misma. Más aún, toda isometría de la esfera puede expresarse como composición de a lo más tres reflexiones.

  Ahora bien, la esfera es una superficie orientable y las reflexiones son isometrías que invierten la orientación, con lo cual las únicas isometrías que conservan la orientación de la esfera son aquellas que pueden expresarse como composición de dos reflexiones. La composición de dos reflexiones en circunferencias maximales es una *rotación* en torno al eje determinado por los dos puntos de corte de las circunferencias maximales dadas. Por otra parte, toda isometría del Plano Proyectivo proviene de dos isometrías de la esfera, siendo cada una de ellas el resultado de componer la otra con la reflexión antipodal. Como la reflexión antipodal es una isometría que invierte la orientación, podemos concluir que una de esas dos isometrías de la esfera conserva la orientación y por lo tanto es una rotación. Por tanto podemos decir que todas las isometrías del Plano Proyectivo son *rotaciones proyectivas*, esto es el paso al cociente de rotaciones de la esfera. Notemos que al ser el Plano Proyectivo una superficie no orientable, no podemos hablar de isometrías que conserven o inviertan su orientación. Una rotación en la esfera deja fijo un par de puntos antipodales y globalmente invariante su circunferencia polar. También quedan invariantes, globalmente, todas las circunferencias centradas en dichos puntos y paralelas a su polar. De aquí deducimos que toda isometría del Plano Proyectivo tiene un punto fijo y deja globalmente invariantes a su recta proyectiva polar y a todas las circunferencias proyectivas centradas en él.

- **Segmento proyectivo**. Teniendo en cuenta que las rectas proyectivas son curvas cerradas homeomorfas a la circunferencia, no tiene sentido hablar de semirrectas proyectivas. Este hecho nos obliga también a ser cuidadosos a la hora de definir segmento proyectivo. En efecto, dos puntos A y B del Plano Proyectivo dividen a la recta proyectiva que determinan en dos componentes cuyas longitudes suman π. Llamaremos *segmento AB* a la componente de menor longitud. En el caso de que ambas componentes midieran $\pi/2$ diremos que el segmento AB no está definido.

- **Distancia Proyectiva**. Si A y B son dos puntos del modelo del disco del Plano Proyectivo sabemos que su distancia proyectiva viene dada por

$$d(A,B) = min[\delta(A_i, B_j)]\ i,j \in 1,2$$

en donde δ denota la distancia esférica y A_1 y A_2 , y B_1 y B_2 son los puntos de la esfera cuyas clases de equivalencia son A y B respectivamente. Como se ha hecho observar anteriormente, a la hora de construir herramientas nos basta con conocer la fórmula de la distancia restringida al caso de que uno de los puntos sea el centro del disco $O = (0,0,0)$. En este caso tenemos:

$$d(O,A) = 2\ \arctan(|z|)$$

en donde z es la coordenada de A visto como punto de la recta compleja.

- **Polígonos**. Para hablar de polígonos en Geometría Proyectiva tenemos que ser un poco cuidadosos al dar la definición de qué entendemos por polígono proyectivo. Dado un número natural $n \geq 3$ diremos que una sucesión de n puntos del Plano Proyectivo A_1, A_2, ..., A_n determinan un *polígono proyectivo* si se verifican las siguientes condiciones:

a) A_{i-1}, A_i y A_{i+1} no están alineados proyectivamente $\forall i : 2 \leq i \leq n-1$.

b) Los segmentos proyectivos A_1A_2, A_2A_3, ...,$A_{n-1}A_n$, A_nA_1 determinan una línea poligonal proyectiva cerrada que no se autointerseca.

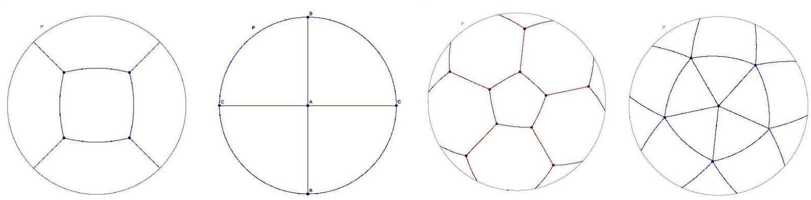
c) La línea poligonal proyectiva anteriormente definida divide al Plano Proyectivo en dos componentes conexas, una de las cuales, a la que llamaremos interior del n-gono proyectivo es homeomorfa a un disco. Nótese que la otra es homeomorfa a una banda de Möbius.

Esta definición de polígono proyectivo nos permite asegurar que toda recta proyectiva que corta a un polígono proyectivo lo hace un número par de veces.

- **Exceso esférico y Teselaciones**. Una propiedad que verifica el Plano Proyectivo es que la suma de los ángulos de cualquier triángulo proyectivo es estrictamente mayor que π y, además, la diferencia entre dicha suma y π aumenta a medida que lo hace el tamaño del triángulo. Llamamos *exceso esférico* a dicha diferencia y nos proporciona la medida de superficie en el Plano Proyectivo. Tenemos así que el área de un triángulo proyectivo de ángulos α, β y γ es igual a $(\alpha + \beta + \gamma) - \pi$.

Nótese que en la Geometría Proyectiva, como en la Hiperbólica, tampoco existe el concepto de semejanza de triángulos ya que si dos triángulos tienen sus ángulos correspondientes iguales entonces tienen la misma área. Nótese asimismo que con ayuda de regla y compás proyectivos podemos construir triángulos equiláteros con lados de cualquier longitud comprendida entre cero y π , cuyos ángulo irán aumentando a medida que lo haga la longitud de sus lados, variando entre $\pi/3$ y $\pi/2$, no alcanzándose el extremo inferior del intervalo. En particular para $n \in \{4, 5\}$ existen triángulos proyectivos equiláteros cuyos ángulos miden exactamente $2\pi/n$. Estos triángulos proyectivos, por sucesivas reflexiones en los lados generan dos *teselaciones* del Plano Proyectivo mediante triángulos proyectivos equiláteros en las que en cada vértice inciden n triángulos, a las que llamaremos respectivamente *tetraedro proyectivo* y *decaedro proyectivo.*

Observemos que el tetraedro proyectivo tiene tres vértices, seis aristas y cuatro caras, mientras que el decaedro proyectivo tiene seis vértices, quince aristas y diez caras. Estas teselaciones son el cociente de las teselaciones inducidas en la esfera por el octaedro y el icosaedro regulares respectivamente.



De modo análogo, a lo que les sucede a los triángulos proyectivos, existen teselaciones por cuadrados proyectivos regulares con números de incidencia en cada vértice igual a 3 y teselaciones por pentágonos proyectivos regulares con números de incidencia en cada vértice igual a 3. A las primeras les llamaremos *triedros proyectivos* y a las segundas *hexaedros proyectivos.*

Observemos que el triedro proyectivo tiene cuatro vértices, seis aristas y tres caras, por lo que es la teselación dual del tetraedro proyectivo, mientras que el hexaedro proyectivo tiene diez vértices, quince aristas y seis caras, por lo que es la teselación dual del decaedro proyectivo. Estas teselaciones son el cociente de las teselaciones inducidas en la esfera por el hexaedro y el dodecaedro regulares respectivamente.

Hagamos notar por último que no todo polígono ni toda teselación de la esfera dan lugar a polígonos y teselaciones del Plano Proyectivo. Así por ejemplo un triángulo equilátero esférico cuyos lados miden más que $\pi/2$ y cuyos ángulos miden, por tanto, también más que $\pi/2$ no da lugar a un triángulo proyectivo, ya que la línea poligonal proyectiva que determinan las proyecciones de sus vértices no divide al Plano Proyectivo. De aquí deducimos que la teselación inducida en la esfera por el tetraedro regular no se proyecta en ninguna teselación del Plano Proyectivo. Tampoco los diedros que teselan la esfera inducen teselación alguna en el Plano Proyectivo ya que sus aristas miden $\pi > \pi/2$ y sus dos vértices dan lugar a un único punto del Plano Proyectivo.

5. Conclusiones

La selección de temas que hemos presentado pone de manifiesto las analogías y las diferencias entre los modelos de Plano Hiperbólico y Proyectivo. Las principales analogías las podemos cifrar en:

1. Que ambos modelos son conformes, obtenidos mediante proyección estereográfica del hiperboloide, en el caso del plano hiperbólico, y de la esfera, en el del plano proyectivo. Hiperboloide y esfera que están sumergidos en el espacio tridimensional $\mathbb{R}^3$ dotado, respectivamente, de la métrica de Lorentz y de la euclídea, y que son modelos de las dos geometrías cuando se identifican sus puntos antipodales.

2. En ambos modelos las rectas de la geometría son arcos de circunferencia.

3. La expresión de la distancia es similar (en el caso proyectivo con funciones trigonométricas y en el hiperbólico con funciones hiperbólicas).

La mayor diferencia entre los modelos de ambas geometrías es la pertenencia, en el caso proyectivo, o no, en el hiperbólico, de la circunferencia borde al modelo. Esta diferencia, debida a la construcción de los modelos, concuerda con el hecho de que el plano proyectivo es acotado y todas las rectas tienen longitud finita, a diferencia de lo que ocurre en el plano hiperbólico.

El modelo proyectivo que hemos construido permite además visualizar otras propiedades de modo sencillo, como son el hecho de que dos rectas siempre se intersequen, la no existencia de semirrectas, la dificultad la definición de segmento y con ello la de polígono, y la construcción de teselaciones, que se dan en el número finito a diferencia de lo que ocurre en el caso hiperbólico.

Por otra parte, existen analogías y diferencias entre las geometrías hiperbólica y proyectiva, que son independientes de los modelos que se consideren de ambas. Ambos son ejemplos de geometrías no euclídeas en los dos modos de negación del quinto postulado de Euclides, y son espacios homogéneos de curvatura constante, en que el exceso esférico y el defecto hiperbólico determinan las áreas de los triángulos. Por contra, el plano proyectivo es no orientable mientras que el hiperbólico sí es orientable, y en el plano proyectivo sólo hay una posición relativa de dos rectas (secantes), mientras que en el hiperbólico hay tres (secantes, paralelas y ultraparalelas). Además existen objetos propios de cada geometría: así los horociclos lo son de la hiperbólica, y la correlación polo -polar lo es de la proyectiva.

Finalmente, como se mencionó al comienzo, ambos modelos permiten su implementación en el programa de Geometría Dinámica GeoGebra, que se convierte así en una útil herramienta para la enseñanza de la Geometría Proyectiva y de la Hiperbólica (véase [2], donde se han empleado las herramientas de Geogebra diseñadas en [3]).

References

[1] D. A. Brannan, M. F. Esplen, J. J. Gray: *Geometry*. Cambridge University Press, 1999.

[2] N. B. Rocha: *Uma proposta de ensino de geometria não-euclidiana utilizando ambiente de geometria dinâmica*. Dissertação apresentada ao Mestrado Profissional em Matemática em Rede Nacional PROFMAT da Universidade Estadual do Sudoeste da Bahia UESB, como requisito necessário para a obtenção do grau de Mestre em Matemática, Universidade Estadual do Sudoeste da Bahia, Brasil, 2014.

[3] U. R. Trías: *Herramientas hiperbólicas y proyectivas en GeoGebra*. Trabajo Fin de Máster. Dirigido por Fernando Etayo. Universidad de Cantabria, 2012.
http://repositorio.unican.es/xmlui/bitstream/handle/10902/578/Ujue

Toroidal Dehn Surgeries

Mario Eudave-Muñoz[*] and Enrique Ramírez-Losada

Instituto de Matemáticas
Universidad Nacional Autónoma de México
Ciudad Universitaria, 04510 México, D.F., Mexico,
and CIMAT, A.P. 402, Guanajuato 36000, Gto., México
mario@matem.unam.mx

CIMAT
A.P. 402, Guanajuato 36000, Gto., México
kikis@cimat.mx

Dedicated to José María Montesinos on the occasion of his 70th birthday.

ABSTRACT

We construct an infinite family of hyperbolic knots K_n, such that for a certain slope σ, $K_n(\sigma)$ has an incompressible torus T which intersects the core of the surgered solid torus in 4 points. For a subfamily of these knots, $K_n(\sigma)$ is a graph manifold, consisting of the union of two Seifert fibered spaces over the disk, one with 2 exceptional fibers, the other with 3 exceptional fibers. Other subfamilies contain asymmetric knots and knots with arbitrarily high tunnel number. We also construct hyperbolic knots whose exteriors contain properly embedded essential Klein bottles with arbitrarily large number of boundary components.

2010 Mathematics Subject Classification: 57M25, 57M50.
Key words: Dehn surgery, incompressible torus, Klein bottle, hitting number, JSJ decomposition.

---

[*]Author partially supported by project PAPIIT-UNAM IN102814

1. Introduction

Let K be a knot in S^3, denote by $\eta(K)$ a regular neighborhood of K, and by $E(K)$ the exterior of K, that is, $E(K) = S^3 - \operatorname{int}\eta(K)$. Let r be a slope in $\partial\eta(K)$, i.e., the isotopy class of a simple closed curve in $\partial\eta(K)$; slopes in $\partial\eta(K)$ are parameterized by $\mathbb{Q} \cup \{1/0\}$ as usual. Denote by $K(r)$ the manifold obtained by performing r-Dehn surgery on K, that is, the result of gluing a solid torus and $E(K)$ along their boundaries such that a meridian of the solid torus is identified with a curve of slope r. If K is not a torus knot, and not a satellite knot, i.e., there is no essential torus in $E(K)$, then by results of Thurston, K is a hyperbolic knot and all but finitely many surgeries on K produce hyperbolic manifolds. By the solution of the Geometrization Conjecture by Perelman, it is known that if a surgery is exceptional, i.e., it fails to be hyperbolic, then it produces a reducible manifold, a Seifert fibered space, or a toroidal manifold, that is, a manifold containing an incompressible torus. It has been a topic of interest to study knots with exceptional surgeries. In this paper we will be concerned with toroidal surgeries.

Suppose that K is a hyperbolic knot and that for a slope r, $K(r)$ is a toroidal manifold. If r is not an integer, then all such possible knots and surgeries have been determined [12], [3]. In fact, in this case $r = p/2$, p an integer, $K(r)$ contains a unique incompressible torus T which intersects the core of the surgered solid torus in two points, and $K(r)$ is a graph manifold which is divided by T into two Seifert fibered spaces over the disk with two exceptional fibers. Furthermore, for any such knot $tn(K) = 1$, where $tn(K)$ denotes the tunnel number of K. On the other hand, if r is an integral slope, there are many more known examples and there is not a classification of all knots with such surgeries. Many examples have been constructed of knots with toroidal surgeries, see for example [4], [8], [13], [18], [21], [22], [23], [24]. Following [23], if T is an incompressible torus in $K(r)$, let K^* be the core of the surgered solid torus, and let $|T \cap K^*|$ be the minimal intersection number between K^* and T. Define the hitting number of the pair (K, r), $ht(K, r)$, to be

$$ht(K, r) = min\{\, |T \cap K^*| \, : T \text{ an incompressible torus in } K(r)\}$$

If K is a genus one knot then $ht(K, 0) = 1$, and by [11] it follows that if $K(0)$ contains a non-separating torus then in fact $ht(K, 0) = 1$, i.e., K is a genus one knot. In other words, if $ht(K, 0)$ is odd then it is equal to one, though it seems to be unknown if there exists a knot K, and a non-separating torus in $K(0)$ with $|T \cap K^*| > 1$. Here we are concerned only with separating tori obtained by Dehn surgery. In fact, for most known examples, $ht(K, r) = 2$. The first examples with $ht(K, r) = 4$ were given in [4]; these were constructed with tangles and using the Montesinos trick. More examples of this type were constructed in [22], where the knots are somehow simpler. In both constructions the resulting knots are strongly invertible. In [23], examples are constructed of toroidal surgeries with hitting number 4 such that the knots are non-strongly invertible, but they are periodic of order two. Furthermore, for any knot

K in [4] or [22], $tn(K) = 1$, and for any knot in [23], $tn(K) = 2$. In those cases, there is a unique incompressible torus with hitting number 4, which divides the manifolds into two Seifert fibered space over the disk with two exceptional fibers. Knots with a toroidal surgery and arbitrarily high tunnel number were constructed in [8], all these examples have a unique incompressible torus and $ht(K, r) = 2$. Examples of knots with arbitrarily high hitting number were first given in [18], and then in [24], it is explicitly proved that for any integer n, there are knots K so that $K(r)$ has a unique incompressible torus and $ht(K, r) = 2n$. In this case $K(r)$ is divided by the incompressible torus into two simple manifolds, in fact, two copies of the exterior of the figure eight knot.

This paper arose as an attempt to solve the following questions:

- How many hyperbolic knots are there with a toroidal surgery with $ht(K, r) = 4$?, or in general with $ht(K, r) = 2n$? How complicated can these knots be?
- If $K(r)$ is a toroidal surgery in a hyperbolic knot K, is there a general upper bound on the number of JSJ pieces of $K(r)$? Or well, is there a general upper bound on the number of non-parallel disjoint incompressible tori that can exist in $K(r)$? This question was posed by K. Motegi [25].
- Is there a knot K so that $E(K)$ contains a properly embedded incompressible Klein bottle with 3 or more boundary components? or in other words, is there a knot K such that $K(r)$ contain a Klein bottle B with $|B \cap K^*| \geq 3$. This is relevant in our case, because if a manifold $K(r)$ contains a Klein bottle B, then in most cases $\partial\eta(B)$ will be an incompressible torus.

In this paper we show that there are in fact plenty of hyperbolic knots K so that $ht(K, r) = 4$. Given any non-satellite unknotting number one knot U and any torus knot $T_{p,q}$, we construct an infinite family of hyperbolic knots K_n, such that for a certain slope σ, $K_{\pm 1}(\sigma)$ has a unique incompressible torus and $ht(K_{\pm 1}, \sigma) = 4$, such a torus divides $K_{\pm 1}(\sigma)$ into a manifold homeomorphic to $E(U)$ and a Seifert fibered space over a disk with two exceptional fibers. By suitably choosing U, we get asymmetric knots and knots with arbitrarily high tunnel number, having a toroidal surgery with hitting number 4. But if $n \neq \pm 1$, then $K_n(\sigma)$ is toroidal, it has a torus T with hitting number 4, such that T divides $K_n(\sigma)$ into a manifold homeomorphic to $E(U)$ and a Seifert fibered space over a disk with three exceptional fibers. As a consequence, $K_n(\sigma)$ contains infinitely many tori, with arbitrarily high hitting numbers. However $ht(K_n, \sigma) = 2$, for there is a torus disjoint from T with hitting number 2. In particular if U is the trefoil knot, we get knots K_n such that $K_n(\sigma)$ is a graph manifold, made of two JSJ pieces, both being a Seifert fibered space over a disk, one with two exceptional fibers, and the other with 3 exceptional fibers. This seems to be the first given example of this type. This is done in Section 2.

In Section 3, we show that there is a family of hyperbolic knots K_n, so that for each knot in the family and any positive odd number m, there is a punctured Klein bottle properly embedded in $E(K_n)$ whose boundary consists of m components of a certain slope r, that is, there are infinitely many Klein bottles in $E(K_n)$ with the same slope. However, we have not produced an example of a knot which has just one Klein bottle with more than one boundary component. For these knots, $K_n(r)$ is toroidal, it has 3 disjoint, incompressible, non-parallel tori, of hitting number 2, but the JSJ decomposition has 3 pieces, two being the exterior of 2-bridge knots, and the third a Seifert fibered space over a once punctured Möbius band with none or just one exceptional fiber. Then $K_n(r)$ contains infinitely many incompressible Klein bottles and tori.

2. Examples of toroidal surgeries with hitting number 4

To construct the examples, we proceed as in [6]. First we construct a 4-punctured torus S properly embedded in the exterior of a knot k, by adding tubes to a pair of annuli, such that S is incompressible towards one side. The surface S is compressible towards the other side, then we find a knot L in its complement so that S is incompressible in $E(k \cup L)$, but the point is that L has to be a trivial knot. Then by doing $1/n$-Dehn surgery on L, we get the desired family of knots K_n.

Let T be a standard torus in S^3. Consider a product neighborhood $T \times I$ of T in S^3, $I = [1,2]$, and identify T with $T \times \{3/2\}$. Let R_1 and R_2 be the complementary solid tori, where $\partial R_i = T \times \{i\}$, $i = 1,2$. Let k be a (p,q)-torus knot emdedded in $T \times \{3/2\}$, where we assume that $p, q \neq 0, \pm 1$, that is, k is a nontrivial knot. Let $\eta(k)$ be a regular neighborhood of k, assume that it is contained in $T \times I$, and that $\eta(k) = A \times I$, where $A \subset T$ is an annulus; hence $A \times \{3/2\}$ is a regular neighborhood of k in $T \times \{3/2\}$. Let $A_1 = cl((T - A) \times \{1\})$ and $A_2 = cl((T - A) \times \{2\})$ be essential annuli properly embedded in $E(k)$. Let C be a disk contained in $T - A$, and let $B = C \times I$, this is a 3-ball; let $C_1 = C \times \{1\} = B \cap A_1$ and $C_2 = C \times \{2\} = B \cap A_2$. Consider a 2-string tangle (B, t_1, t_2), where each t_i is an arc properly embedded in B with one endpoint in C_1 and the other one in C_2. Note that this tangle has a natural framing, that is, a choice of meridian and latitude, given by ∂C and the product structure on $\partial C \times I$.

Assume that the tangle (B, t_1, t_2) has the following properties:

1. The closure of (B, t_1, t_2) shown in the left of Figure 1 is the trivial knot.
2. The numerator U of (B, t_1, t_2), that is, the closure of (B, t_1, t_2) shown in the right of Figure 1 is not the trivial knot.
3. If Q is the exterior of the tangle, that is $Q = B - int(\eta(t_1) \cup \eta(t_2))$, where $\eta(t_1)$ and $\eta(t_2)$ are disjoint regular neighborhoods of t_1 and t_2 in B, then Q is atoroidal.

Note that (1) implies that (B, t_1, t_2) is not a product tangle, that is, t_1, t_2 are not a pair of monotonic arcs in the product $C \times I$. It follows that the numerator U

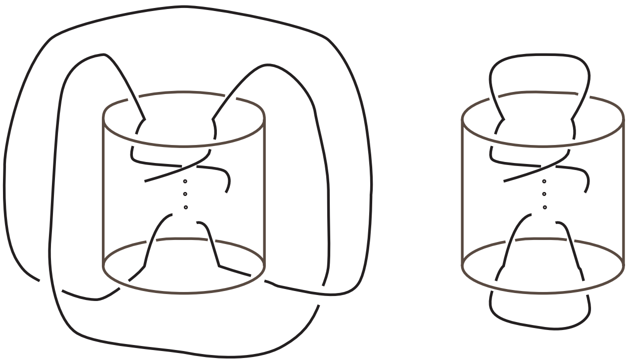


Figure 1: The tangle (B, t_1, t_2) and two of its closures

of the tangle is a nontrivial knot with unknotting number one. In fact, it is a non-satellite unknotting number one knot, for if there is an essential torus in $E(U)$, then by [19] it can be made disjoint from any unknotting crossing, which would contradict that Q is atoroidal; in particular, U is a prime knot. Now, if U is a non-satellite unknotting number one knot, then by taking the complement of a ball enclosing a crossing change, we get a tangle (B, t_1, t_2) with the required properties, for if there is an essential torus in the exterior of (B, t_1, t_2), the torus will be essential in $E(U)$.

Let $G_i = cl(\partial\eta(t_i) - \partial B)$, $i = 1, 2$; note that G_i is an annulus. Let $\hat{A}_i = cl(A_i - (\eta(t_1) \cup \eta(t_2)))$, $i = 1, 2$; this is a twice punctured annulus. Finally let $P = \hat{A}_1 \cup G_1 \cup G_2 \cup \hat{A}_2$, that is, we are joining the punctured annuli $\hat{A}_1$ and $\hat{A}_2$ with the tubes G_1 and G_2. It follows that P is a punctured torus properly embedded in $E(k)$, whose boundary consists of 4 curves of slope pq.

Let L_i, $i = 1, 2$, be an arc in $T \times \{i\}$ joining the endpoints of t_1 and t_2, such that L_i intersects a meridian of R_i in one point and it is disjoint from a preferred longitud of R_i. Let L be the knot $L = t_1 \cup L_1 \cup t_2 \cup L_2$. There are two choices for each of the arcs L_i, but with loss of generality, assume that L is the closure of the tangle (B, t_1, t_2) shown in the left of Figure 1. Now isotope L such that it lies in the complement of k and P. Then L consists of the arcs t_1, t_2 plus two arcs L_1, L_2, contained in R_1, R_2, and which go around the cores of R_1 and R_2 respectively. Note that L is independent of the torus knot k, that is, it just depends on $T \times I$ and the tangle (B, t_1, t_2). By our previous assumption, L is the trivial knot.

In Figure 2 we show the simplest example, where we consider the $(3, 2)$-torus knot, and (B, t_1, t_2) is the rational tangle $R(-3)$. For rational tangle we follow the convention of [5].

So, P is a 4-punctured torus properly embedded in $E(k \cup L)$, the exterior of $k \cup L$, but that ∂P is contained in $\partial\eta(k)$, and then P is disjoint from L. Note that P divides $E(k \cup L)$, into two pieces, V and W, where, say, $\partial N(L)$ lies in V. Let $E_i = (\eta(t_i) \cap A_1) - \operatorname{int}\eta(L)$, $i = 1, 2$; so this is an annulus properly embedded in

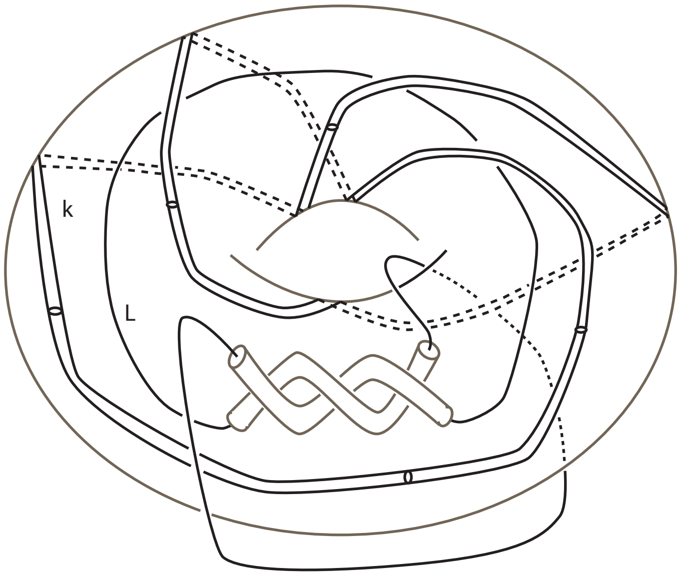


Figure 2: A link that produces a family of knots having a toroidal surgery with hitting number 4.

V, one boundary component lying in P and the other one in $\partial N(L)$; note that E_i is incompressible in V. ∂P divides $\partial\eta(k)$ in four annuli, which we denote F_1, F_2, F_3, F_4, where, say, F_1 and F_2 lie in V, while F_3 and F_4 lie in W.

Let K_n be the knot obtained from k after performing $1/n$-Dehn surgery on L, $n \in \mathbb{Z}\backslash\{0\}$, and let P_n be the image of the torus P. Let σ be the slope of ∂P_n in the exterior of K_n, that is, the image of the slope pq in k after performing $1/n$-surgery on L. Note that $lk(k, L) = |p \pm q|$, hence $\sigma = pq - n|p \pm q|^2$.

Let $K_n(\sigma)$ be the manifold obtained by σ-surgery on K_n, and let $\hat{P}_n$ be the torus in $K_n(\sigma)$ obtained by capping off P_n with meridian disks of the surgered solid torus. The torus $\hat{P}_n$ divides $K_n(\sigma)$ in two parts, denoted by $\hat{V}$ and $\hat{W}$, which come from V and W respectively. Note that $\hat{W}$ is obtained from W by adding two 2-handles along the annuli F_3 and F_4. It is not difficult to see that $\hat{W}$ is homeomorphic to $E(U)$, the exterior of the numerator of the tangle (B, t_1, t_2). Note also the $\hat{V}$ is obtained from V by performing $1/n$-surgery on L and then adding two 2-handles along the annuli F_1, F_2. Note that V can be seen as the solid tori R_0, R_1 joined by a pair of 1-handles, from which we have removed a neighborhood of L. By adding a 2-handle to R_1 along F_1, we get a punctured lens space $L(p, q)$, and by adding a 2-handle to R_2 along F_2, we get a punctured lens space $L(q, p)$. So by reversing the order of the surgery and the handle additions, we can see $\hat{V}$ as obtained by first joining the punctured lens spaces $L(p, q)$ and $L(q, p)$ with two 1-handles, getting a manifold M, then taking a knot L in M which goes around each 1-handle once, and which goes once around the core of each lens spaces, and finally doing $1/n$-surgery on L.

Lemma 2.1 *$\hat{V} \cong D^2(a/p, b/q, c/n)$, i.e., $\hat{V}$ is a Seifert fibered space over the disk with two exceptional fibers of indices p and q, and a third exceptional fiber of index n if $n \neq \pm 1$. If $n = \pm 1$, $\hat{V}$ has just two exceptional fibers.*

Proof. Consider the pieces in which $M - \operatorname{int} \eta(L)$ is divided by the annuli E_1 and E_2, which we denote by M_1 and M_2. We want to show that M_1 and M_2 are solid tori. Note that the manifold M_1 comes from the solid torus R_1, from which we have deleted the neighborhood of the arc $L \cap R_1$, and then glued a 2-handle along a curve of slope (p, q) on ∂R_1, i.e., along the core of the annulus F_1. The manifold $R_1 - \operatorname{int} \eta(L \cap R_1)$ is homeomorphic to the manifold obtained from R_1 by deleting an open neighborhood of the 1-complex $\tau_1 \cup \tau_2$, formed by the core τ_1 of R_1 and a straight arc τ_2 joining τ_1 and ∂R_1 (this can be seen just by sliding τ_1 over τ_2, for we get the arc $L \cap R_1$). There is an spanning annulus A in the product $R_1 - \operatorname{int} \eta(\tau_1)$, where one of its boundary components lie in ∂R_1 and is disjoint from the annulus F_1, and the other lies in $\partial\eta(\tau_1)$. A boundary component of A is then a curve of slope (p, q) on ∂R_1. It follows that the annulus A intersects a meridian of L in p points. We can assume that the arc τ_2 lie in A, so that after removing an open neighborhood of τ_2, $A - \operatorname{int} \eta(\tau_2)$ is a disk in M_1, that is, is a compression disk for M_1 which intersects each meridian of L in p points, and which is disjoint from F_1, and then disjoint from the 2-handle attached to R_1 along F_1. By compressing $R_1 - \operatorname{int} \eta(L \cap R_1)$ with this disk, we get a solid torus, whose longitude is the image of the curve (p, q), and then, after gluing the 2-handle along F_1 we get a 3-ball. That is, we are showing that M_1 is a solid torus, where one of its meridians intersect a meridian of the knot L in p points. So, M_1 has a Seifert fibration, with fibers parallel to meridians of L, that is, fibers parallel to ∂E_1, and an exceptional fiber of index p. Similarly, M_2 has a Seifert fibration, with fibers parallel to meridians of L, and an exceptional fiber of index q. Note that M_1 and M_2 are then glued along two annuli made of fibers to get M. So, M is a Seifert fibered space over an annulus with two exceptional fibers of indices p and q. Finally, by doing $1/n$-Dehn surgery on L, we get the desired result. □

Lemma 2.2 *The JSJ decomposition of $K_n(\sigma)$ is just $\hat{V} \cup \hat{W}$, and the torus $\hat{P}_n$ is the unique torus in this decomposition. It is a graph manifold only when the knot U is the trefoil knot.*

Proof. First, it is clear that $\hat{P}_n$ is an incompressible torus in $K_n(\sigma)$. Note that $\hat{W}$ is a Seifert fibered space only when $\hat{W}$ is homeomorphic to the exterior of a torus knot, but the trefoil knot is the only torus knot with unknotting number one. Note that in this case, a fiber of $\hat{V}$ is identified with a meridian of $\hat{W}$. □

If the knot U is chosen to be the trivial knot, the above proof shows that $K_n(\sigma)$ is a Seifert fibered space over the sphere with three (or two) exceptional fibers. There are two possibilities for the tangle (B, t_1, t_2) such that U is the trivial knot. This produces two possibilities for the trivial knot L. In this case the knot L is what is called a seiferter, and these particular seiferters for torus knots are studied in detail in [1].

Note that if $n = \pm 1$, then $\hat{P}_n$ is the unique incompressible torus in $K_n(\sigma)$. But if $n \neq \pm 1$, then in fact there are infinitely many torus in $K_n(\sigma)$, all contained in $\hat{V}$, as this is a Seifert fibered space with 3 exceptional fibers. Infinitely many of these tori can be described as follows, where M_1 and M_2 are as in the proof of Lemma 2.1. Consider the torus S_1 formed by the union of $\partial M_1 \cap \partial \hat{V}$, the annuli E_1 and E_2, and the annulus $\partial \eta(L) \cap M_2$, and similarly the torus S_2 formed by the union of $\partial M_2 \cap \partial \hat{V}$, the annuli E_1 and E_2, and the annulus $\partial \eta(L) \cap M_1$. After an isotopy, we can assume that these tori intersect in two curves γ_1 and γ_2, and that both are disjoint from $\hat{P}_n$. After doing a Dehn twist along S_1, the torus S_2 transforms into a torus S_2^1. In general, after m Dehn twists, the torus S_2 transforms into a torus S_2^m. Similarly, after m Dehn twists along the torus S_2, the torus S_1 transforms into a torus S_1^m.

Denote by K_n^* the core of the surgered solid torus. In $\hat{V}$ and $\hat{W}$, K_n^* can be seen as the core of the attached 2-handles. Note also that $\hat{V}$ with the two arcs of K_n^* removed, is the same as V after Dehn surgery on L, which is a compression body, so it cannot contain an incompressible torus, that is, any incompressible torus in $\hat{V}$ intersects K_n^*.

Lemma 2.3 *Any incompressible torus in $K_n(\sigma)$ intersects K_n^*. If $n = \pm 1$, $ht(K_{\pm 1}, \sigma) = 4$. If $n \neq \pm 1$, $ht(K_n, \sigma) = 2$.*

Proof. Note that any torus in $K_n(\sigma)$ is either the torus $\hat{P}_n$, or it is contained in the interior of $\hat{V}$. As noted above, any incompressible torus in $\hat{V}$ intersects K_n^*. It follows that if $n \neq \pm 1$, then $ht(K_n, \sigma) = 2$. This is because $|S_1 \cap K_n^*| = 2$, and there is no incompressible torus disjoint form K_n^*.

Suppose that $n = \pm 1$. Then $\hat{P}_{\pm 1}$ is the unique incompressible torus in $K_{\pm 1}(\sigma)$ and $|\hat{P}_{\pm 1} \cap K_{\pm 1}^*| = 4$, so it just have to be proved that this intersection is minimal. If the intersection is not minimal, by Proposition 1.1 of [4], there is an arc component t of $cl(K_{\pm 1}^* - \hat{P}_{\pm 1})$, which is contained in either $\hat{W}$ or in $\hat{V}$, such that t can be isotoped to lie in a product neighborhood of $\hat{P}_{\pm 1}$. So, after removing a neighborhood of t, the resulting manifold contains a copy of the original manifold in which t lies. If t lies in $\hat{W}$, after removing t we should get a manifold containing an incompressible torus, which is not possible, for what we get after removing t is a copy of Q, the exterior of the tangle (B, t_1, t_2), which is atoroidal. If t lies in $\hat{V}$, then after removing t, we get a genus two handlebody, which then cannot contain a copy of $\hat{V}$. □

With a little more work it can be shown that for any n we have $|\hat{P}_n \cap K_n^*| = 4$, and that for $i = 1, 2$, and for any n and m, $|S_i^m \cap K_n^*| = 2 + 4m$. The above proof also shows that P_n is incompressible in $E(K_n)$.

Lemma 2.4 *K_n is a hyperbolic knot.*

Proof. It follows from Lemma 3.10 of [7]. □

We summarize the previous results.

Theorem 2.1 *Given a torus knot (p, q) and a non-satellite unknotting number one knot U, there is a collection of hyperbolic knots K_n, such that doing σ-Dehn surgery on K_n, by a certain slope σ, produces a toroidal manifold with the following properties:*

1. *If $n \neq \pm 1$, $K_n(\sigma) = E(U) \cup_{\hat{P}_n} D^2(a/p, b/q, c/n)$, where $\hat{P}_n$ is an incompressible torus, and $D^2(a/p, b/q, c/n)$ is a Seifert fibered space over the disk with three exceptional fibers of indices p, q, and n. In this case $ht(K_n, \sigma) = 2$.*

2. *If $n = \pm 1$, $K_n(\sigma) = E(U) \cup_{\hat{P}_n} D^2(a/p, b/q)$, where $\hat{P}_n$ is the unique incompressible torus in $K_n(\sigma)$, and $D^2(a/p, b/q)$ is a Seifert fibered space over the disk with two exceptional fibers of indices p and q. In this case $ht(K_n, \sigma) = 4$.*

3. *If U is the trefoil knot and $n \neq \pm 1$, then $K_n(\sigma)$ is a graph manifold made of two Seifert fibered pieces, one fibered over a disk with two exceptional fibers of indices 2 and 3, the other fibered over a disk with three exceptional fibers of indices p, q and n.*

We note that all previously knows examples of graph manifolds obtained by surgery, are made of Seifert fibered pieces with only two exceptional fibers. See for example [13], where many such examples are constructed.

Now we consider the tunnel number of the knots K_n. If M is a compact 3-manifold, $Hg(M)$ denotes the Heegaard genus of M. It follows $Hg(K_n(\sigma)) \leq tn(K_n) + 1$.

Proposition 2.5 *Let K_n be constructed as above, such that the tangle (B, t_1, t_2) is a rational tangle. Then $tn(K_n) = 1$.*

Proof. Note that if (B, t_1, t_2) is a rational tangle, then it has to be the rational tangle $R((-2n + 1)/n)$, for some $n \neq 0, 1$. Let τ be an arc properly embedded in $E(K_n)$, which is an spanning arc of the annulus A_1. Note that $E(k \cup \tau)$ can be seen as the solid tori R_0, R_1, connected by the 3-ball $B = C \times I$, i.e., $E(k \cup \tau)$ is a genus two handlebody which contains in its interior the knot L. Note that $\partial E(k \cup \tau)$ is compressible in $E(k \cup \tau) - L$ if an only if (B, t_1, t_2) is a rational tangle. In this case, after performing $1/n$-Dehn surgery on L we get another handlebody, showing that $E(K_n \cup \tau)$ is a genus two handlebody, that is, $tn(K_n) = 1$. □

Proposition 2.6 *If U is a 2-bridge knot, or if $n = \pm 1$, and U is a 1-bridge knot in S^3, then $Hg(K_n(\sigma)) = 2$. In all other cases $Hg(K_n(\sigma)) \geq 3$, and then $tn(K_n) \geq 2$.*

Proof. By the main result of [14], if $Hg(K_n(\sigma)) = 2$, then there are two possibilities for this manifold. First, $Hg(K_n(\sigma)) = D(i) \cup M_K$, where $D(i)$ is a Seifert fibered space over a disk with i exceptional fibers, $i = 2, 3$, and M_K is the exterior of a

2-bridge knot in S^3, such that a fiber of $D(i)$ is glued to a meridian of M_k. Second, $Hg(K_n(\sigma)) = D(2) \cup L_K$, where $D(2)$ is a Seifert fibered space over a disk with 2 exceptional fibers, L_K is the exterior of a non-satellite 1-bridge knot in a lens space, such that a fiber of $D(2)$ is glued to a meridian of L_k. But in our case, a fiber of $D(2)$ has to be glued to a meridian of $E(U)$, so that $E(U)$ must be in fact the exterior of a 1-bridge knot in S^3. □

If $Hg(K_n(\sigma)) = 2$, then $tn(K_n)$ could be 1. However, we do not know if this is true in general, we only know it for the case of Proposition 2.5.

Proposition 2.7 *Let U and K_n be as above, and let $k = tn(U)$.*

1. *If $n = \pm 1$, then $\frac{1}{3}(k+1) \leq Hg(K_n(\sigma)) \leq k+4$, so $\frac{1}{3}(k-2) \leq tn(K_n)$.*
2. *If $|n| \geq 2$, then $\frac{1}{3}(k+2) \leq Hg(K_n(\sigma)) \leq k+5$, so $\frac{1}{3}(k-1) \leq tn(K_n)$.*

Proof. As, $K_n(\sigma) = E(U) \cup \hat{V}$, and $Hg(U) = k+1$, if follows from Lemma 3.5 of [8] that $Hg(E(U)) + Hg(\hat{V}) \leq 3Hg(K_n(\sigma)) + 2$, that is, $\frac{1}{3}(k + Hg(\hat{V}) - 1) \leq Hg(K_n(\sigma))$ (a similar estimate can be obtained from [20]). On the other hand, it is not difficult to show that $Hg(K_n(\sigma)) \leq Hg(E(U)) + Hg(\hat{V}) + 1 = k + Hg(\hat{V}) + 2$. Now note that $Hg(\hat{V}) = 2$ if $n = \pm 1$, and $Hg(\hat{V}) = 3$ if $|n| \geq 2$. □

We now show that there are hyperbolic knots with high tunnel number having a toroidal surgery with hitting number 4. It was known previously that there are hyperbolic knots with high tunnel number having a toroidal surgery [8], but in those examples the hitting number is 2.

Lemma 2.8 *Given any positive integer m, there is an hyperbolic knot U, with unknotting number one and $tn(U) \geq m$.*

Proof. By [26], there is a 2-string tangle whose exterior Q is simple, that is, it is irreducible, ∂-irreducible, anannular and atoroidal. Take N copies $Q_1, Q_2, \dots Q_N$ of the exterior of this tangle, as in Figure 3, shown there for $N = 4$, and let U be a knot in the complement of the Q_i's, with a configuration as in Figure 3. Note that U has unknotting number one. For each Q_i there are two disks $D_{i,1}$ and $D_{i,2}$ whose boundary is essential in Q_i and such that U intersects each $D_{i,j}$ in two points. Let $c = \partial D_{1,1}$, note that $lk(c, U) = 0$. We first show that $c \cup L$ is not a split link. Let B' be a small 3-ball around the clasp of U, i.e., around the crossing that unknots U. By performing the crossing change we get the trivial knot U', and in fact $c \cup U'$ is a split link. Then by [19], if $c \cup L$ is a split link, there is a sphere separating c and U disjoint from the 3-ball B'. Let now U'' be a knot obtained by a tangle replacement in B', then $c \cup U''$ must be a split link, but U'' can be chosen such that $lk(c, U'') = 2$, which is a contradiction. Suppose that some ∂Q_i is compressible in the complement of U, and let D be a compression disk. An standard argument shows that D can be made disjoint from the collection of disks $D_{i,j}$. Now, it is not difficult to see that there is no compression disk disjoint from the $D_{i,j}$, so each ∂Q_i is incompressible.

Suppose that Σ is an essential torus in the complement of U. Because the manifolds Q_i are simple, the torus Σ can be made disjoint from each Q_i, so suppose Σ is contained in the complement of the Q_i's. Note that the complement of the Q_i's, union the disks $D_{i,j}$ and the knot U is a handlebody, so Σ must intersect the disks $D_{i,j}$. Look at the intersection between the torus Σ and the collection of disks $D_{i,j}$. A standard argument shows that the intersection consists of curves that are essential in Σ, and that are parallel to $\partial D_{i,j}$ in each disk $D_{i,j}$. If there are two or more curves in a same disk $D_{i,j}$, by an interchange of annuli we get another essential torus, also denoted Σ, with fewer intersections with the $D_{i,j}$'s. So, Σ intersects each $D_{i,j}$ in at most one curve. Note also that Σ is disjoint from the 3-ball B'. It follows that Σ is decomposed into annuli that connect pair of disks, and that must cross through all the disks $D_{i,j}$. In particular, there is an annulus connecting $D_{1,n-1}$ and $D_{1,n}$ that determines a 3-ball B_1 which contains B', and other annuli connecting $D_{2,n-1}$ and $D_{1,n}$, that determines a 3-ball B_2, such that $B_1 \cap B_2 \subset D_{1,n}$. Connect the endpoints of U lying in D_{2n-1}, and the endpoints lying in D_{1n}. By looking at the Figure 3, we get a curve c' which is isotopic to c. On the other hand, c' lies in the 3-ball B_2, and then is a trivial curve in the complement of U, which contradicts that c' is isotopic to c.

It follows that given any integer N, there exists a hyperbolic knot K with unknotting number one such that its exterior contains $\geq N$ incompressible, non-parallel surfaces of genus 2. By the main result of [9], there is a constant $C(2,t)$, such that if K is a knot with tunnel number t, then the exterior of K contains at most $C(2,t)$ disjoint, nonparallel, incompressible surfaces of genus 2. So, if we choose $N > C(2, m-1)$, it follows that $tn(U) \geq m$.

A better estimate for the number of copies Q_i that are needed could be obtained from the main result of [20], but to use that, we have to find a bound for the number of annuli contained in the complement of the Q_i's. $\square$

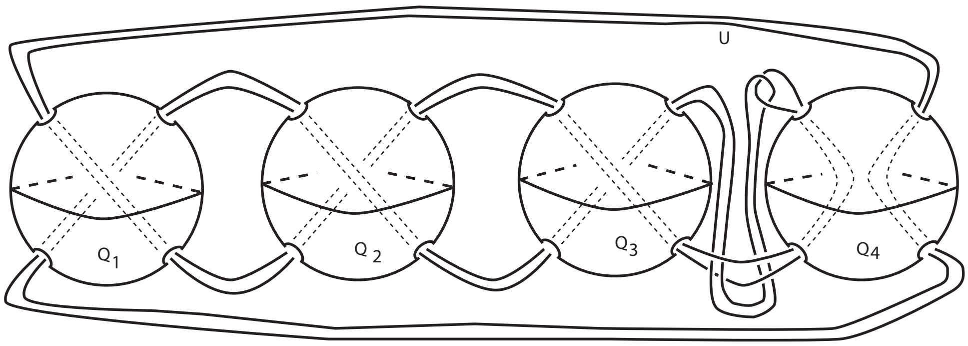


Figure 3: A knot with unknotting number one and high tunnel number.

Theorem 2.2 *Given any positive integer m, there is a hyperbolic knot K with a toroidal surgery of hitting number 4, such that $tn(K) \geq m$.*

Proof. Choose a hyperbolic unknotting number one knot U with $tn(U) \geq 3m + 2$, and let $K = K_{\pm 1}$. By Proposition 2.7 it follows that $tn(K) \geq m$. □

Let K be a knot in S^3. A symmetry of K is an orientation preserving PL homeomorphism $f : S^3 \to S^3$ of finite order, such that $h(K) = K$. All known examples of knots with toroidal surgeries of hitting number 4 have a symmetry. Most are strongly invertible, and in [23] examples are constructed which are non-strongly invertible, but they are periodic of order two. Here we show the following result.

Theorem 2.3 *There are asymmetric hyperbolic knots with a toroidal surgery of hitting number 4.*

Proof. Let U be an asymmetric, hyperbolic knot with unknotting number one. Knots of this type exists, take for example the knots 9_{33} and 10_{82} [2]. Consider the knot $K_{\pm 1}$ constructed in Theorem 2.1 by using U and any torus knot. Let h be a periodic homeomorphism of $K_{\pm 1}$. The map h extends to a periodic homeomorphism of $K_{\pm 1}(\sigma)$. By the $\mathbb{Z}_p$-Equivariant Torus Theorem [16], and because there is a unique essential torus $\hat{P}_{\pm 1}$ in $K_{\pm 1}(\sigma)$, it follows that $\hat{P}_{\pm 1}$ is invariant under the action of h. Such a map h cannot interchange the complementary regions of $\hat{P}_{\pm 1}$. So, the exterior of U admits a non-trivial periodic homeomorphism, and then U admits a symmetry, a contradiction. □

As noted in Proposition 2.5, some of the knots K_n have tunnel number one and hence are strongly invertible. So, the toroidal surgeries on these knots can also be explained by using tangles and double branched covers. Here, we just describe one example. In Figure 4 we have a knot $\mathcal{K}_n$, formed by the union of two tangles along a sphere S, one of them being a Montesinos tangle of length 3, the other a Montesinos tangle of length 2. We show a 3-ball B intersecting $\mathcal{K}_n$ in two arcs; note that B intersects S in two disks. By doing the tangle replacement indicated in B, that is, changing two vertical arcs with two horizontal arcs with the crossings indicated, we get the trivial knot. Let K_n be the knot obtained from the link $k \cup L$ of Figure 2 by performing $1/n$-surgery on L. The exterior of the knot K_n double branch covers the tangle determined by the complement of B, $K_n(\sigma)$ double branch covers $\mathcal{K}_n$ and the torus $\hat{P}_n$ double branch covers the sphere S.

3. Examples of knots with many Klein bottles

To construct the examples, we first construct a branched surface $\mathcal{B}$ by adding tubes to 3 disks embedded in the exterior of the trivial knot K, then find a trivial knot in the complement of the surface, such that $\mathcal{B}$ is incompressible in $E(K \cup L)$. Again, by doing $1/n$-Dehn surgery on L, we get the desired family on knots K_n. Such a branched surface will carry infinitely many Klein bottles and tori.

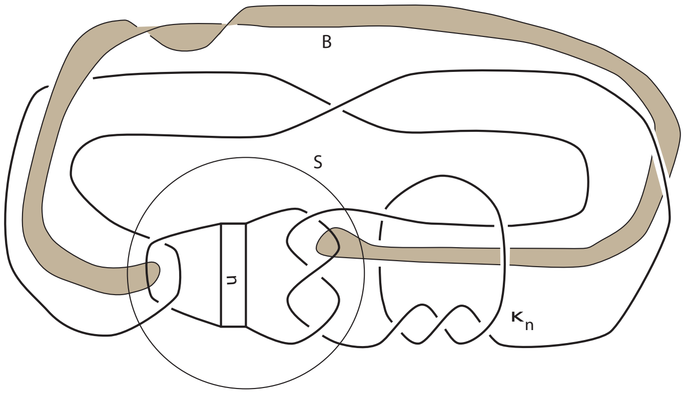


Figure 4: A knot made of Montesinos tangles of length 2 and 3 which can be unknotted by a band move.

Let $B = D\times[1,3]$, where D is a disk and $D_i = D\times\{i\}$, $i = 1, 2, 3$. Let (B, τ_1, τ_2) be the tangle shown in the left of Figure 2, so τ_1, τ_2 are a pair of arcs properly embedded in B, with its endpoints lying in $D_1 \cup D_3$. In this Figure, an horizontal or vertical box denotes a string of $\pm p$ horizontal crossings or q vertical crossings, respectively. Assume that p is odd and $|p| \geq 3$, and that q is even, $|q| \geq 2$. Note that (B, τ_1, τ_2) is a partial sum of two rational tangles B_1 and B_2, which are glued along the disk D_2. The disk D_2 intersects both string of the tangle, because p is odd and q is even.

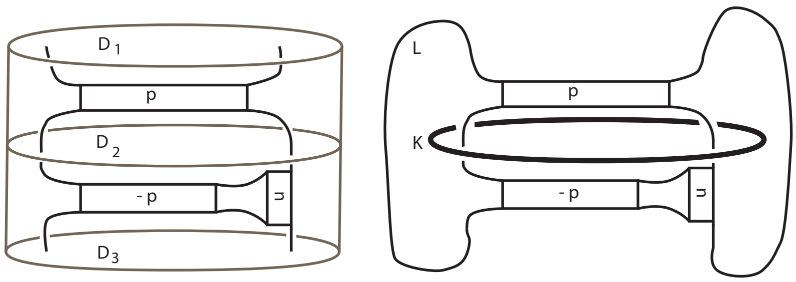


Figure 5: The tangle (B, τ_1, τ_2) and its denominator.

Assume (B, τ_1, τ_2) is embedded in S^3, just as in the left of Figure 5. Let L be the denominator of (B, τ_1, τ_2), it is the trivial knot. Let K be the boundary of the disk D_2, as in the right of Figure 5. Assume that $\eta(K)$ is a regular neighborhood of K such that $\eta(K) \cap B = \partial D \times [1, 3]$, and such that $\eta(K)$ and L are disjoint. Then, the disks D_1, D_2 and D_3 are properly embedded in $E(K)$. In Figure 6 we are showing the case $p = 3$, $n = 2$. Consider the link $K \cup L$, and note that the linking number

$lk(K, L) = \pm 2$.

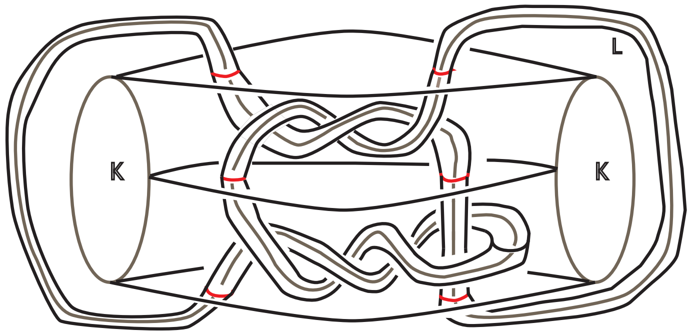


Figure 6: A 2-complex which produces a branched surface.

Each of the disks D_i intersects L in two points. So, the knot L intersects the collection of disks D_i in 6 points, which then divide L in 6 arcs, denoted by $t_1, t_2, t_3, t_4, t_5, t_6$, where t_1, t_4 go from D_1 to D_2, t_2, t_5 go from D_2 to D_3, and t_3, t_6 go from D_3 to D_1. Let $\eta(L)$ be a regular neighborhood of L, which can be expressed as a union of neighborhoods $\eta(t_i)$ of the arcs t_i. Let $A_i = cl(\partial\eta(t_i) - (\cup D_j))$. Let $\hat{D}_i = cl(D_i - \cup\eta(t_j))$. Note that ∂D_i consists of 3 simple closed curves, one is a longitude of the knot K, denote the other two boundary curves by c_{i1} and c_{i2}. We assume that the annulus A_6 connects c_{11} and c_{31} and that A_1 connects c_{11} with c_{22}. It follows that A_4 connects c_{12} with c_{21}, A_2 connects c_{22} with c_{32}, and A_5 connects c_{21} with c_{31}, see Figure 6. Consider the 2-complex $W = \hat{D}_1 \cup A_1 \cup A_4 \cup \hat{D}_2 \cup A_2 \cup A_5 \cup \hat{D}_3 \cup A_6$, that is, consider the union of the 3 punctured disks plus 5 of the annuli (all, except A_3). This complex is a surface except along 4 simple closed curves, namely c_{11}, c_{21} c_{22} and c_{31}. Smooth the intersections of A_1 with $\hat{D}_1$ and $\hat{D}_2$, and the intersections of A_5 with $\hat{D}_2$ and $\hat{D}_3$, to get a branched surface $\mathcal{B}$ properly embedded in $E(K \cup L)$ as in Figure 7. In the figure we are showing the case $p = 3$ and $n = 2$, after doing a reduction in the number of crossings of L. We refer to [10] for facts about branched surfaces. The singular locus of the branched surface consists of 4 simple closed curves, again the curves c_{11}, c_{21} c_{22} and c_{31}. Note that $\mathcal{B} - \{\text{singular curves}\}$ consists of 6 components, denoted A, B, C, D, E, F, where $A = \hat{D}_1 \cup A_4$, $B = \hat{D}_2$, $C = \hat{D}_3 \cup A_2$, $D = A_1$, $E = A_5$ and $F = A_6$. Let N be a fibered neighborhood of $\mathcal{B}$, and let $\partial_h N$ the horizontal boundary of N. Note that each complementary region of the singular curves have two copies in $\partial_h N$. If S is such a region other than B, let S_- denote the copy that is closer to L, that is, the one that contains a curve that cobounds with a meridian of L an annulus with interior disjoint from N, and let S_+ be the other copy. For the piece B, let B_- be the copy lying between D_2 and D_3, and B_+ be the copy lying between D_1 and D_2. Note that $\partial_h N$ has four components, namely, F_+, $B_+ \cup D_+ \cup A_+ - c_{21}$, $B_- \cup E_+ \cup C_+ - c_{2,2}$, $C_- \cup D_- \cup F_- \cup E_- \cup A_- - c_{1,1}$.

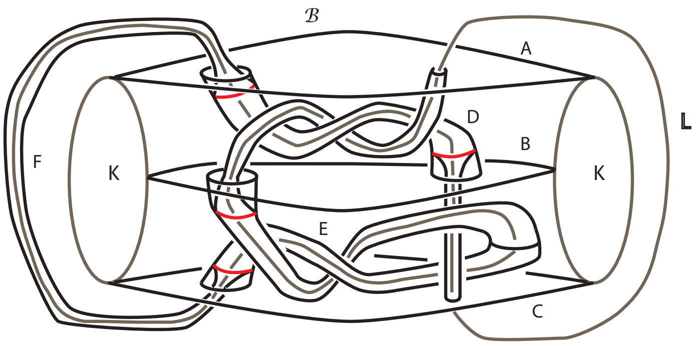


Figure 7: A branched surface in the exterior of the link $K \cup L$.

Let K_n be the knot obtained after performing $1/n$-Dehn surgery on L, $n \in \mathbb{Z}\backslash\{0\}$, ?and let $\mathcal{B}_n$ be the branched surface in $E(K_n)$ which is the image of $\mathcal{B}$ after Dehn surgery on L. The slope of the boundary of $\mathcal{B}$ is just the 0 slope, so the slope of $\mathcal{B}_n$ is $-4n$, because $lk(K,L) = 2$.

Lemma 3.1 *$\mathcal{B}$ is an incompressible branched surface in $E(K \cup L)$.*

Proof. By inspection, it is not difficult to see that there are no disks or half-disks of contact in N. Now we show that each component of $\partial_h N$ is incompressible and ∂-incompressible in $E(K \cup L) - \operatorname{int} N$. Consider first the surface $B_+ \cup D_+ \cup A_+ - c_{21}$ and let Δ be a compression disk. If $\partial\Delta$ is a non-separating curve in this surface, Δ will be also a compression disk for the exterior of the knot obtained as numerator of the tangle B_1, which is a nontrivial knot. If $\partial\Delta$ is a separating curve, by cutting the surface along Δ we get either a disk and a once punctured torus, or an annulus and a torus, so we get either a disk with slope 0 or a torus embedded in the exterior of B_1. The first case is not possible, for $lk(K,L) = 2$. If we get a torus, then it must be compressible, for the exterior of B_1 is a handlebody, and then a compression disk for the torus will be also a compression disk for $B_+ \cup D_+ \cup A_+ - c_{21}$ along a non-separating curve. The same proof shows that there are no monogons in $E(K \cup L) - \operatorname{int} N$ consisting of an arc in $B_+ \cup D_+ \cup A_+ - c_{21}$ and one arc in a neighborhood of c_{21}.

The proof for the component $B_- \cup E_+ \cup C_+ - c_{2,2}$ is similar. For the component $C_- \cup D_- \cup F_- \cup E_- \cup A_- - c_{1,1}$, just note that there is an annulus E_1 in $E(K \cup L) - \operatorname{int} N$, one boundary being a meridian of L, the other a curve on A_2. If Δ is a compression disk, by eliminating intersections with E_1, we get that $\partial\Delta$ is contained in A or C, but this is not possible. Note also that the component F_+ is clearly incompressible. It follows from [10] that $\mathcal{B}$ is incompressible. □

Lemma 3.2 *Suppose that $n \neq \pm 1$. Then $\mathcal{B}_n$ is an incompressible branched surface in $E(K_n)$.*

Proof. Let N be a fibered neighborhood of $\mathcal{B}_n$. Again by inspection, there are no disks or half-disks of contact in N and there are no monogons in $E(K \cup L) - \operatorname{int} N$. To see that each piece of $\partial_h N$ is incompressible and ∂-incompressible in $E(K_n)$, do an argument similar to that of Theorem 4 of [15], or that of Theorem 2.4 of [6]. This argument fails for $n = \pm 1$, and in fact it can be seen that the piece $C_- \cup D_- \cup F_- \cup E_- \cup A_- - c_{1,1}$ compress after ± 1-Dehn surgery on L. □

Lemma 3.3 *$K \cup L$ is an hyperbolic link.*

Proof. Let Q be the exterior of the tangle (B, τ_1, τ_2). Note that ∂Q is a genus 2 surface which can be pushed to lie in the interior of $E(K \cup L)$. By [26], Q is a simple manifold. Note that there are two nonparallel annuli E_1, E_2 in the complement of Q, one of its boundary components is a meridian of L, the other is an essential curve on ∂Q. By a simple innermost disk-outermost arc argument between these disks and a potential compression disk, it can be shown that ∂Q is in fact incompressible in $E(K \cup L)$. Let now be Σ an essential torus in $E(K \cup L)$. Σ can be isotoped to be disjoint from Q, just because Q is simple. Then Σ is contained in the complement of Q, and by looking again at the intersections between Σ and the disks E_1, E_2, we get that Σ is compressible or peripheral, hence non-essential. A similar argument shows that there are no essential annuli in $E(K \cup L)$. □

Lemma 3.4 *K_n is a hyperbolic knot.*

Proof. Let ∂Q_n be the image of ∂Q after Dehn surgery on L. Again, by Theorem 4 of [15], or Theorem 2.4 of [6], it follows that ∂Q_n in incompressible in $E(K_n)$. If Σ is a torus in the complement of Q_n, do an argument as in Theorem 4 of [15], looking at the intersections between Σ and the annulus E_1. □

Let G^s, $s \geq 1$, be the surface carried by $\mathcal{B}_n$ and given by the weights:
$a = s, \quad b = 1, \quad c = s, \quad d = s + 1, \quad e = s + 1, \quad f = 1.$

A weight $a = s$, it just means that we are taking s copies of the branch A of $\mathcal{B}$. It is not difficult to see that G^n is a $(2n + 1)$-punctured Klein bottle. In Figure 8, we show the surface G^3.

Let F^r, $r \geq 1$, be the surface carried by $\mathcal{B}_n$ and given by the weights:
$a = r, \quad b = 2, \quad c = r, \quad d = r + 2, \quad e = r + 2, \quad f = 2.$

It is not difficult to see that F^r is an $(2r + 2)$-punctured tori. Note also that if $r = 2s$, then F^r is in fact the double of G^s.

Theorem 3.1 *Suppose that $n \neq \pm 1$. The surfaces G^s and F^r are incompressible and ∂-incompressible in $E(K_n)$.*

Proof. This is just a direct consequence of Theorem 2 of [10]. □

The knots K_n have more interesting properties.

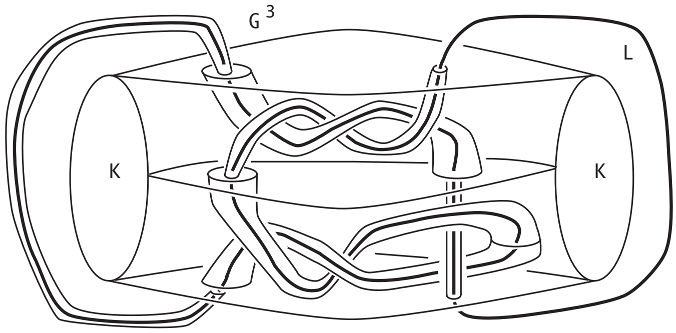


Figure 8: A Klein bottle with 3 boundary components.

Proposition 3.5 *Suppose that $n \neq \pm 1$. There is an incompressible and ∂-incompressible once-punctured Klein bottle properly embedded in $E(K_n)$, which is carried by $\mathcal{B}_n$.*

Proof. Take the surface carried by $\mathcal{B}_n$ with weights $a = 0$, $b = 1$, $c = 0$, $d = 1$, $e = 1$, $f = 1$. It is not difficult to see that it is a once-punctured Klein bottle G^1. Note that the branched surface does not contain Reeb componentes, and then by Theorem 2 of [17], it follows that G^1 is incompressible and ∂-incompressible. □

Theorem 3.2 *There are 3 disjoint incompressible 2-punctured tori properly embedded in $E(K_n)$, with the same slope as $\mathcal{B}_n$.*

Proof. Consider the tori $T_1 = \hat{D}_1 \cup A_1 \cup A_4 \cup \hat{D}_2$, $T_2 = \hat{D}_2 \cup A_2 \cup A_5 \cup \hat{D}_3$, $T_3 = \hat{D}_1 \cup A_1 \cup A_4 \cup A_2 \cup A_5 \cup \hat{D}_3$. By an isotopy, these tori can be made disjoint. A similar proof as in Lemma 3.1 show that these tori are incompressible in $E(K \cup L)$. By an application of Theorem 4 of [15], it follows that the tori are incompressible in $E(K_n)$. □

Now consider $-4n$-Dehn surgery on K_n, that is, surgery along the slope of the boundary of $\mathcal{B}$. We can interchange the order of the surgeries and the result is the same, so first do 0-surgery on K and then $1/n$-Dehn surgery on L. By doing 0-surgery on K, we get $S^1 \times S^2$. By capping off the boundary components of $\mathcal{B}$ with meridian disks of the glued solid tori, we get a branched surface $\tilde{\mathcal{B}}$, as shown in Figure 9, where the two spheres are being identified. By the same argument as before, it follows that $\tilde{\mathcal{B}}$ is an incompressible branched surface in $S^1 \times S^2 - \operatorname{int} \eta(L)$, which remains incompressible after $1/n$-Dehn surgery on L, $n \neq \pm 1$, i.e., $K_n(-4n)$ contains an incompressible branched surface $\tilde{\mathcal{B}}_n$. So, $K_n(-4n)$ carries infinitely many incompressible Klein bottle and tori, carried by $\tilde{\mathcal{B}}_n$, which come by capping off the surfaces G^s, F^r.

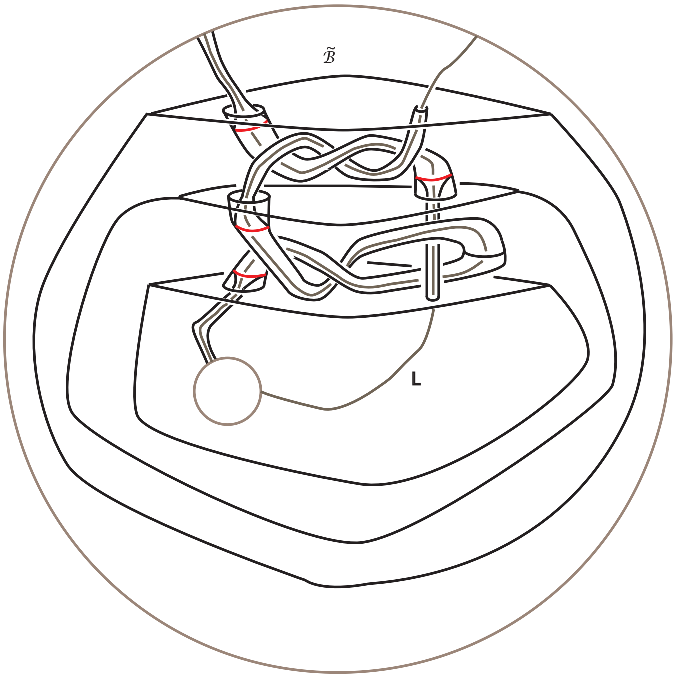


Figure 9: A branched surface in $S^11 \times S^2$.

Consider the tori $\tilde{T}_1$, $\tilde{T}_2$, $\tilde{T}_3$, obtained by capping off the tori T_1, T_2, T_3. Again, these tori are disjoint and incompressible in $K_n(\sigma)$, for any n. The existence of these tori implies that $ht(K_n, -4n) = 2$.

Theorem 3.3 *The JSJ decomposition of $K_n(-4n)$ is given by the tori $\tilde{T}_1$ and $\tilde{T}_2$, they separate $K_n(-4n)$ in 3 pieces, two are exteriors of 2-bridge knots and one is a Seifert fibered space over a once punctured Möbius band with no exceptional fibers if $n = \pm 1$, and one exceptional fiber of index n if $n \neq \pm 1$.*

Proof. As said above, the tori $\tilde{T}_1$, $\tilde{T}_2$, $\tilde{T}_3$ are incompressible in $K_n(-4n)$. Note that $\tilde{T}_2$ bounds a 3-manifold homeomorphic to the exterior of the 2-bridge knot $T(p, n)$, which is simple. The torus $\tilde{T}_1$ bounds a 3-manifold homeomorphic to the exterior of the torus knot $T(2, n)$, which is then a Seifert fibered space. The tori $\tilde{T}_1$, $\tilde{T}_2$, $\tilde{T}_3$ cobound a manifold homeomorphic to $(pair\, of\, pants) \times S^1$, which is also Seifert fibered space, but its fibers are identified to meridians of the knot exteriors bounded by $\tilde{T}_1$ and $\tilde{T}_2$, so the fibers in both pieces do not coincide. The torus $\tilde{T}_3$ bounds a manifold M, as shown in Figure 10, where the two interior spheres are being identified. This manifold can be fibered with fibers parallel to meridians of the knot L, the result being a Seifert fibered space over a once punctured Möbius band with no exceptional fibers. Note that the fibers of M and the fibers of $(pair\,of\,pants) \times S^1$ coincide, then its union is a Seifert fibered space over a twice punctured Möbius band with no exceptional fibers. Finally, by doing Dehn surgery on L, we get the desired result. □

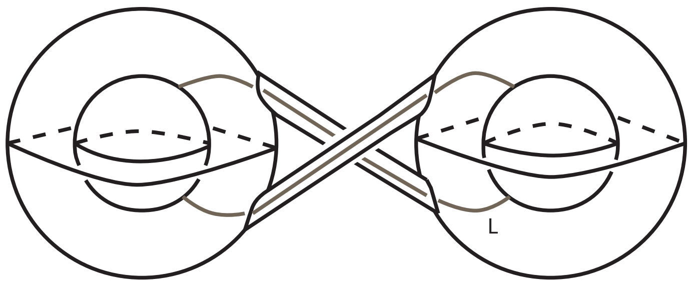


Figure 10: A Seifert fibered space over a once punctured Möbius band.

Proposition 3.6 $tn(K_n) = 2$.

Proof. By the main result of [14], $K_n(-4n)$ cannot have a genus two Heegaard splitting, so $Hg(K_n(-4n)) \geq 3$, and then $tn(K_n \geq 2$. Now, it is not difficult to check, by explicitly finding the tunnels, that $tn(K \cup L) = 2$, and then $tn(K_n) = 2$. □

Finally note that a little more general construction can be made just by considering sums of rational tangles whose denominator is a trivial knot, and the results obtained will be the same.

References

[1] A. Deruelle, K. Miyazaki, K. Motegi: *Networking Seifert surgeries on knots, III.* Algebr. Geom. Topol. **14**, 2065–2101 (2014).

[2] J. C. Cha, C. Livingston: *Table of Knot Invariants.* http://www.indiana.edu/∼linkinfo, May 27, 2015.

[3] M. Eudave-Muñoz: *Non-hyperbolic manifolds obtained by Dehn surgery on hyperbolic knots.* In *Geometric Topology. 1993 Georgia international topology conference,* W. H. Kazez (ed.), August 2-13, 1993, Athens, GA, USA. AMS/IP Stud. Adv. Math. 2, 35–61, AMS, Providence, RI, 1997.

[4] M. Eudave-Muñoz: *4-punctured tori in the exteriors of knots.* J. Knot Theory Ramifications **6**, 659–676 (1997).

[5] M. Eudave-Muñoz: *On hyperbolic knots with Seifert fibered Dehn surgeries.* Topology Appl. **121**, 119–141 (2002).

[6] M. Eudave-Muñoz: *On knots with icon surfaces.* Osaka J. Math. **50**, 271–285 (2013).

[7] M. Eudave-Muñoz, E. Jasso, K. Miyazaki, K. Motegi: *Seifert fibered surgeries on strongly invertible knots without primitive/Seifert positions.* Topology Appl., to appear.

[8] M. Eudave-Muñoz, J. Luecke: *Knots with bounded cusp volume yet large tunnel number.* J. Knot Theory Ramifications **8**, 437–446 (1999).

[9] M. Eudave-Muñoz, J. Shor: *A universal bound for surfaces in 3-manifolds with a given Heegaard genus.* Algebr. Geom. Topol. **1**, 31–37 (2001).

[10] W. Floyd, U. Oertel: *Incompressible surfaces via branched surfaces.* Topology **23**, 117–125 (1984).

[11] D. Gabai: *Foliations and the topology of 3-manifolds, III.* J. Differ. Geom. **26**, 479–536 (1987).

[12] C.McA. Gordon, J. Luecke: *Non-integral toroidal Dehn surgeries.* Commun. Anal. Geom. **12**, 417–485 (2004).

[13] S. Kang: *Knots in S^3 admitting graph manifold Dehn surgeries.* J. Korean Math. Soc. **51**, 1221–1250 (2014).

[14] T. Kobayashi: *Structures of the Haken manifolds with Heegaard splittings of genus two.* Osaka J. Math. **21**, 437–455 (1984).

[15] W. Menasco, *Closed incompressible surfaces in alternating knot and link complements.* Topology **23**, 37–44 (1984).

[16] W. Meeks, P. Scott, *Finite group actions on 3-manifolds.* Invent. Math. **86**, 287–346 (1986).

[17] U. Oertel: *Incompressible branched surfaces.* Invent. Math. **76**, 385–410 (1984).

[18] J. Osoinach: *Manifolds obtained by surgery on an infinite number of knots in S^3.* Topology **45**, 725–733 (2006).

[19] M. Scharlemann, A. Thompson: *Unknotting number, genus and companion tori.* Math. Ann. **280**, 191–205 (1988).

[20] J. Schultens: *Heegaard genus formula for Haken manifolds.* Geom. Dedicata **119**, 49–68 (2006).

[21] M. Teragaito: *Toroidal surgeries on hyperbolic knots.* Proc. Am. Math. Soc. **130**, 2803–2808 (2002).

[22] M. Teragaito: *Hyperbolic knots with three toroidal Dehn surgeries.* J. Knot Theory Ramifications **17**, 1051–1061 (2008).

[23] M. Teragaito: *Non-invertible knots having toroidal Dehn surgery of hitting number four.* Hiroshima Math. J. **38**, 447–454 (2008).

[24] M. Teragaito: *Toroidal Dehn surgery on hyperbolic knots and hitting number.* Topology Appl. **157**, 269–273 (2010).

[25] Y. Tsutsumi: *Universal bounds for genus one Seifert surfaces for hyperbolic knots and surgeries with non-trivial JSJT-decompositions.* Interdiscip. Inf. Sci. **9**, 53–60 (2003).

[26] Y-Q. Wu: *The classification of nonsimple algebraic tangles.* Math. Ann. **304**, 457–480 (1996).

The open quadrant problem: A topological proof

José F. Fernando[*], **J.M. Gamboa**[†] and **Carlos Ueno**[‡]

Departamento de Álgebra
Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
28040 Madrid, Spain
josefer@mat.ucm.es jmgamboa@mat.ucm.es

Dipartimento di Matematica
Università degli studi di Pisa
56127 Pisa, Italy
cuenjac@mail.dm.unipi.it

Dedicated to José María Montesinos on the occasion of his 70th birthday.

ABSTRACT

In this work we present a new polynomial map $f := (f_1, f_2) : \mathbb{R}^2 \to \mathbb{R}^2$ whose image is the open quadrant $\mathcal{Q} := \{x > 0, y > 0\} \subset \mathbb{R}^2$. The proof of this fact involves arguments of topological nature that avoid hard computer calculations. In addition each polynomial $f_i \in \mathbb{R}[\mathtt{x}, \mathtt{y}]$ has degree ≤ 16 and only 11 monomials, becoming the simplest known map solving the open quadrant problem.

2010 Mathematics Subject Classification: 14P10, 26C99, 52A10.
Key words: Polynomial map, polynomial image, semialgebraic set, open quadrant, total degree, total number of monomials.

[*]Author supported by Spanish GAAR MTM2011-22435, Grupos UCM 910444 and the "National Group for Algebraic and Geometric Structures, and their Applications" (GNASA - INdAM). His one year research stay in the Dipartimento di Matematica of the Università di Pisa is partially supported by MECD grant PRX14/00016.

[†]Author supported by Spanish GAAR MTM2011-22435 and Grupos UCM 910444.

[‡]Author supported by 'Scuola Galileo Galilei' Research Grant at the Dipartimento di Matematica of the Università di Pisa and Spanish GAAR MTM2011-22435.

1. Introduction

Although it is usually said that the first work in Real Geometry is due to Harnack [13], who obtained an upper bound for the number of connected components of a non-singular real algebraic curve in terms of its genus, modern Real Algebraic Geometry was born with Tarski's article [15], where it is proved that the image of a semialgebraic set under a polynomial map is a semialgebraic set. We are interested in studying what might be called the 'inverse problem'. In the 1990 *Oberwolfach Reelle algebraische Geometrie* week [12] the second author proposed:

Problem 1.1 *Characterize the (semialgebraic) subsets of* $\mathbb{R}^m$ *that are either polynomial or regular images of* $\mathbb{R}^n$.

A map $f := (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ is a *polynomial map* if its components $f_k \in \mathbb{R}[\mathtt{x}] := \mathbb{R}[\mathtt{x}_1, \ldots, \mathtt{x}_n]$ are polynomials. Analogously, f is a *regular map* if its components can be represented as quotients $f_k = \frac{g_k}{h_k}$ of two polynomials $g_k, h_k \in \mathbb{R}[\mathtt{x}]$ such that h_k never vanishes on $\mathbb{R}^n$. A subset $S \subset \mathbb{R}^n$ is *semialgebraic* when it admits a description by a finite boolean combination of polynomial equalities and inequalities.

Open semialgebraic sets deserve a special attention in connection with the real Jacobian Conjecture [14]. In particular the second author stated in [12] the 'open quadrant problem':

Problem 1.2 *Determine whether the open quadrant* $\mathcal{Q} := \{x > 0, y > 0\}$ *of* $\mathbb{R}^2$ *is a polynomial image of* $\mathbb{R}^2$.

This problem stimulated the interest of many specialists in the field. However, only after twelve years a first solution was found in [4] and presented by the first author in the 2002 *Oberwolfach Reelle algebraische Geometrie* week [2].

The open quadrant problem was the germ of a more systematic study of 'Polynomial and regular images of Euclidean spaces' developed by the authors during the last decade and which was the topic of the Ph.D. Thesis of the third author [16]. Since then we have worked on this issue with two main objectives:

- Finding obstructions to be an either polynomial or regular image.
- Proving (constructively) that large families of semialgebraic sets with piecewise linear boundary (convex polyhedra, their interiors, complements and the interiors of their complements) are either polynomial or regular images of some Euclidean space. The positive answer to the open quadrant problem has been a recurrent starting point for this approach.

In [4, 5] we presented the first steps to approach Problem 1.1. A complete solution to Problem 1.1 for the one-dimensional case appears in [3], whereas in [6, 8, 9, 17, 18] we approached constructive results concerning the representation as either polynomial or regular images of the semialgebraic sets with piecewise linear boundary commented above. Articles [7, 10] are of different nature because we find in them new obstructions for a subset of $\mathbb{R}^m$ to be either a polynomial or a regular image of $\mathbb{R}^n$. In the first one we found some properties of the difference $\mathrm{Cl}(S) \setminus S$ while in the second it is shown that *the set of points at infinite of a polynomial image of* $\mathbb{R}^n$ *is a connected set.*

The constructive solution to the open quadrant problem provided in [4] involves quite complicated computer calculations that the third author never liked. In fact he provided in his Ph.D. Thesis a different topological proof for the map proposed in [4], together with an algebraic proof involving a different polynomial map. This map has inspired the first and third authors for a short algebraic proof of the open quadrant problem involving a new polynomial map [11] and has led us to look for a polynomial map with optimal algebraic structure whose image is the open quadrant. It is important to establish clearly the meaning of 'optimal algebraic structure' [11, §3(A)]. It is natural to wonder how a polynomial map looks like when completely expanded and how it compares with other polynomial maps. We care about the total degree of the involved polynomial map (the sum of the degrees of its components) and its total number of (non-zero) monomials. We would like to find a polynomial map with the less possible total degree and the less possible number of monomials. The example in [4] has total degree 56 and its total number of monomials is 168. The polynomial map in [11] has total degree 72 and its total number of monomials is 350. In this work we will prove:

Theorem 1.3 *The open quadrant* $\mathcal{Q}$ *is the image of the polynomial map*

$$f : \mathbb{R}^2 \to \mathbb{R}^2,\ (x, y) \mapsto ((x^2y^4 + x^4y^2 - y^2 - 1)^2 + x^6y^4, (x^6y^2 + x^2y^2 - x^2 - 1)^2 + x^6y^4).$$

This polynomial map has total degree 28 and its total number of monomials is 22, which certainly improves the already known explicit solutions to the open quadrant problem. It has been constructed following a similar strategy to that in [4, §3]. Our experience approaching this problem suggests us that this map is surely close to have the optimal desired algebraic structure.

The article is organized as follows. In Section 2 we present all basic notions and topological preliminaries used in Section 3 to prove Theorem 1.3.

2. Topological preliminaries

Denote the closed disc of center the origin and radius $A > 0$ of the plane $\mathbb{R}^2$ with $\mathbb{D}_A$. A *warped disc* is a subset $\mathcal{D}_{A,\xi} := \{z = \xi(x, y),\ x^2+y^2 \leq A^2\} \subset \mathbb{R}^3$ where $\xi : \mathbb{R}^2 \to \mathbb{R}$ is a continuous function. Consider the homeomorphism

$$\zeta : \mathbb{R}^3 \to \mathbb{R}^3,\ (x, y, z) \mapsto (x, y, z - \xi(x, y))$$

that maps $\mathcal{D}_{A,\xi}$ onto $\mathbb{D}_A \times \{0\}$. The image of $\mathcal{D}_{A,\xi}$ under a permutation of the variables of $\mathbb{R}^3$ will be also called a warped disc.

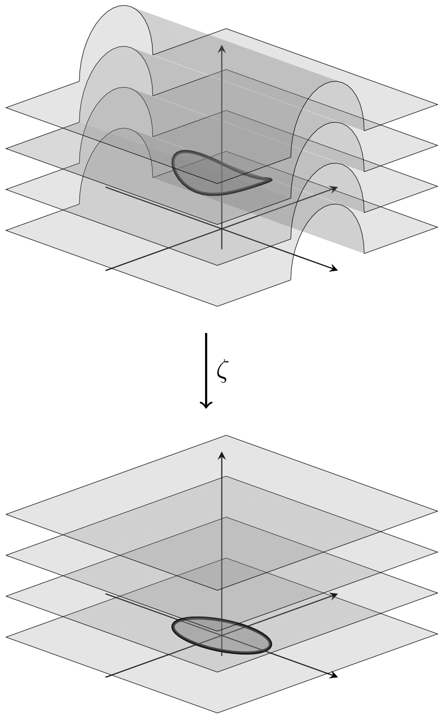


Figure 1: The homeomorphism ζ for $\xi(x,y) := \sqrt{B^2 - \min(y^2, B^2)}$ acting on $\mathbb{R}^3$.

For each $\varepsilon > 0$ consider the open neighborhood

$$\mathbb{D}_A(\varepsilon) := \{x^2 + y^2 < (A+\varepsilon)^2\} \times (-\varepsilon, \varepsilon) \subset \mathbb{R}^3$$

of $\mathbb{D}_A$. Clearly, $\mathcal{D}_{A,\xi}(\varepsilon) := \zeta^{-1}(\mathbb{D}_A(\varepsilon))$ is an open neighborhood of $\mathcal{D}_{A,\xi}$ in $\mathbb{R}^3$.

Definition 2.1 A (continuous) path $\alpha : [a,b] \to \mathbb{R}^3$ *meets transversally once the warped disc* $\mathcal{D}_{A,\xi}$ if there exist $s_0 \in (a,b)$ and $\varepsilon > 0$ such that $J := \alpha^{-1}(\mathcal{D}_{A,\xi}(\varepsilon)) = (s_0 - \varepsilon, s_0 + \varepsilon)$ is an open subinterval of $[a,b]$ and $(\zeta \circ \alpha)|_J(t) = (0, 0, t - s_0)$.

Remark 2.2 If the path $\alpha : [a,b] \to \mathbb{R}^3$ meets transversally once the warped disc $\mathcal{D}_{A,\xi}$, then $\alpha([a,b]) \cap \partial\mathcal{D}_{A,\xi} = \varnothing$.

Let C be a topological space homeomorphic to a closed disc and let $\phi : C \to \mathbb{R}^3$ be a continuous map. The restriction $\partial\phi := \phi|_{\partial C}$ is called the *boundary map* of ϕ. We say that the boundary map $\partial\phi$ *meets transversally once* a warped disc $\mathcal{D}_{A,\xi} \subset \mathbb{R}^3$ if there exists a parameterization β of ∂C such that $\alpha := \phi \circ \beta$ meets transversally once the warped disc $\mathcal{D}_{A,\xi}$.

Given a path-connected topological space X and a point $x_0 \in X$ we denote the fundamental group of X at the base point x_0 with $\pi_1(X, x_0)$. Each path α starting and ending at x_0 is called a loop with base point x_0 and represents an element of $\pi_1(X, x_0)$, that we denote with $[\alpha]$.

Lemma 2.3 *Let $\mathcal{D}_{A,\xi}$ be a warped disc of $\mathbb{R}^3$ and let $X := \mathbb{R}^3 \setminus \partial\mathcal{D}_{A,\xi}$. Let $\alpha : [a, b] \to X$ be a loop with base point $x_0 \in X$ that meets transversally once $\mathcal{D}_{A,\xi}$. Then $[\alpha]$ is a generator of $\pi_1(X, x_0) \cong \mathbb{Z}$.*

Proof. Keep the notations introduced above. Let $s_0 \in (a, b)$ and $\varepsilon > 0$ be such that

$$J := \alpha^{-1}(\mathcal{D}_{A,\xi}(\varepsilon)) = (s_0 - \varepsilon, s_0 + \varepsilon)$$

is an open subinterval of $[a, b]$ and $(\zeta \circ \alpha)|_J(t) = (0, 0, t - s_0)$. After a reparameterization of α we may assume $s_0 = 0$.

As ζ is a homeomorphism of $\mathbb{R}^3$, we will prove the statement for $\beta := \zeta \circ \alpha$, $Y := \mathbb{R}^3 \setminus \partial\mathbb{D}_A$ and the base point $y_0 := \beta(-\varepsilon) = (0, 0, -\varepsilon)$. Consider the path $\gamma : [0, 1] \to \mathbb{R}^3$ given by

$$\gamma(t) := \begin{cases} (3(A+\varepsilon)t, 0, \varepsilon) & \text{if } 0 \le t \le \frac{1}{3}, \\ (A+\varepsilon, 0, \varepsilon - (t - \frac{1}{3})6\varepsilon) & \text{if } \frac{1}{3} < t \le \frac{2}{3}, \\ (A+\varepsilon - 3(A+\varepsilon)(t - \frac{2}{3}), 0, -\varepsilon) & \text{if } \frac{2}{3} < t \le 1. \end{cases}$$

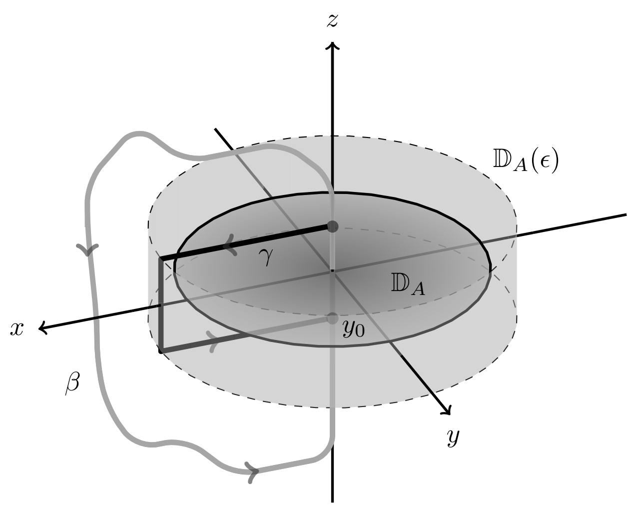


Figure 2: The path β meets transversally once the disk $\mathbb{D}_A$.

Write $\beta_0 := \beta|_J$ and $\beta_1 := \beta|_{[\varepsilon,b]} * \beta|_{[a,-\varepsilon]}$. We claim:

$$[\beta] = [\beta_0 * \beta_1] = [\beta_0 * \gamma] \cdot [\gamma^{-1} * \beta_1] = g \cdot e = g,$$

where e and g are respectively the identity element and a generator of $\pi_1(Y, y_0) \cong \mathbb{Z}$.

The loop $\gamma^{-1} * \beta_1$ with base point y_0 is contained in $\mathbb{R}^3 \setminus \mathbb{D}_A$, which is a simply connected space. Consequently, $[\gamma^{-1} * \beta_1] = e$ in $\pi_1(Y, y_0)$.

The class $[\beta_0 * \gamma]$ generates $\pi_1(Y, y_0)$. Indeed, Y has as deformation retract the set $Z := \partial\mathbb{D}_A(\varepsilon) \cup I_\varepsilon$ where $I_\varepsilon := \{(0,0)\} \times \{-\varepsilon \le z \le \varepsilon\}$. It is an exercise of algebraic topology to show that $[\beta_0 * \gamma]$ is a generator of $\pi_1(Z, y_0) \cong \pi_1(Y, y_0) \cong \mathbb{Z}$, as required. $\square$

Lemma 2.4 *Let $\phi : C \to X$ be a continuous map and assume that C is homeomorphic to a closed disc. Let $\beta : [a, b] \to \partial C$ be a parameterization starting and ending at $z_0 \in \partial C$. Then $[\phi \circ \beta]$ is the identity element of $\pi_1(X, \phi(z_0))$.*

Proof. Let $\psi : C \to \{x^2 + y^2 \le 1\}$ be a homeomorphism. The continuous map

$$H : [0, 1] \times [a, b] \to X, \ (\rho, t) \mapsto (\phi \circ \psi^{-1})(\rho \cdot (\psi \circ \beta)(t) + (1 - \rho) \cdot \psi(z_0)))$$

is a homotopy map between $\phi \circ \beta$ and the constant path, as required. $\square$

Proposition 2.5 *Let C be a topological space homeomorphic to a closed disc and $\phi : C \to \mathbb{R}^3$ a continuous map. Assume $\partial\phi : \partial C \to \mathbb{R}^3$ meets transversally once a warped disc $\mathcal{D} \subset \mathbb{R}^3$. Then $\partial\mathcal{D} \cap \phi(\operatorname{Int}(C)) \neq \varnothing$.*

Proof. Assume by contradiction $\partial\mathcal{D} \cap \phi(\operatorname{Int}(C)) = \varnothing$. As $\partial\phi$ meets transversally once $\mathcal{D}$, the image $\phi(\partial C)$ does not intersect $\partial\mathcal{D}$ by Remark 2.2. Thus, $\phi(C) \subset X := \mathbb{R}^3 \setminus \partial\mathcal{D}$. Let $\beta : [a, b] \to \partial C$ be a parameterization starting and ending at $z_0 \in \partial C$ such that $\phi \circ \beta$ meets transversally once $\mathcal{D}$. By Lemma 2.4 the class $[\phi \circ \beta]$ is the identity element of $\pi_1(X, \phi(z_0))$. However, by Lemma 2.3 the class $[\phi \circ \beta]$ is a generator of $\pi_1(X, \phi(z_0)) \cong \mathbb{Z}$, which is a contradiction. Consequently, $\partial\mathcal{D} \cap \phi(\operatorname{Int}(C)) \neq \varnothing$, as required. $\square$

3. Proof of Theorem 1.3

Observe first that the map f in the statement of Theorem 1.3 is the composition $f_2 \circ f_1$ of the polynomial maps

$$f_1 : \mathbb{R}^2 \to \mathbb{R}^2, \quad (x, y) \mapsto (x^2, y^2),$$
$$f_2 : \mathbb{R}^2 \to \mathbb{R}^2, \quad (x, y) \mapsto ((xy^2 + x^2y - y - 1)^2 + x^3y^2, (x^3y + xy - x - 1)^2 + x^3y^2).$$

As $f_1(\mathbb{R}^2)$ is the closed quadrant $\overline{\mathcal{Q}} := \{x \ge 0, y \ge 0\}$, we have to prove the equality

$$f_2(\overline{\mathcal{Q}}) = \mathcal{Q}. \tag{3.1}$$

The inclusion $f_2(\overline{\mathcal{Q}}) \subset \mathcal{Q}$ is straightforward because both components of f_2 are strictly positive on $\overline{\mathcal{Q}}$. It only remains to show the inclusion

$$\mathcal{Q} \subset f_2(\overline{\mathcal{Q}}). \tag{3.2}$$

3.1. Reduction of the proof of inclusion (3.2)

Consider the (continuous) semialgebraic maps

$$g : \overline{\mathcal{Q}} \to \mathbb{R}^3,\ (x,y) \mapsto (xy^2 + x^2y - y - 1, x^{3/2}y, x^3y + xy - x - 1)$$
$$h : \mathbb{R}^3 \to \mathbb{R}^2,\ (x,y,z) \mapsto (x^2 + y^2, y^2 + z^2).$$

As $f_2 = h \circ g$, we have to show that for each tuple $(A^2, B^2) \in \mathcal{Q}$ there exists $(x_0, y_0) \in \overline{\mathcal{Q}}$ such that $(h \circ g)(x_0, y_0) = (A^2, B^2)$. This is equivalent to check that the intersection $h^{-1}(\{(A^2, B^2)\}) \cap g(\overline{\mathcal{Q}})$ is non-empty.

Denote $\mathcal{S} := g(\overline{\mathcal{Q}})$ and fix values $B \geq A > 0$. It holds that sets

$$h^{-1}(\{(A^2, B^2)\}) = \{x^2 + y^2 = A^2, y^2 + z^2 = B^2\},$$
$$h^{-1}(\{(B^2, A^2)\}) = \{y^2 + z^2 = A^2, x^2 + y^2 = B^2\}$$

contain respectively the boundaries of the warped discs

$$\mathcal{D}_1 : z = \xi_1(x,y), \quad x^2 + y^2 \leq A^2, \tag{3.3}$$
$$\mathcal{D}_2 : x = \xi_2(y,z), \quad y^2 + z^2 \leq A^2, \tag{3.4}$$

for the (continuous) semialgebraic functions

$$\xi_1 : \mathbb{R}^2 \to \mathbb{R},\ (x,y) \mapsto \sqrt{B^2 - \min\{y^2, B^2\}}, \tag{3.5}$$
$$\xi_2 : \mathbb{R}^2 \to \mathbb{R},\ (y,z) \mapsto \sqrt{B^2 - \min\{y^2, B^2\}}. \tag{3.6}$$

Consequently, we are reduced to prove:

3.1.1. *For fixed values $B \geq A > 0$ the intersections $\partial\mathcal{D}_1 \cap \mathcal{S}$ and $\partial\mathcal{D}_2 \cap \mathcal{S}$ are non-empty.*

3.2. Proof of Statement 3.1.1

Write $\mathcal{R} := [0, +\infty) \times (0, \frac{\pi}{2})$ and $\overline{\mathcal{R}} := [0, +\infty) \times [0, \frac{\pi}{2}]$. Consider the map $\phi := (\phi_1, \phi_2, \phi_3) : \mathbb{R}^2 \to \mathbb{R}^3$ where

$$\begin{aligned}
\phi_1(\rho, \theta) &:= \cos\theta \sin\theta(\cos\theta - \sin\theta)^2 \\
&\qquad + \rho(2\cos^4\theta \sin\theta + \cos\theta \sin^4\theta + \cos^5\theta) + \rho^2 \cos^5\theta \sin\theta, \\
\phi_2(\rho, \theta) &:= \sqrt{\cos\theta \sin\theta}(\cos\theta + \sin\theta + \rho\cos\theta\sin\theta), \\
\phi_3(\rho, \theta) &:= \rho \sin\theta.
\end{aligned}$$

Let us prove now some properties of the map ϕ and the sets $\mathcal{R}$ and $\overline{\mathcal{R}}$:

3.2.1. $\phi(\mathcal{R}) \subset \mathcal{S}$.

Proof. The analytic map

$$\psi : \mathcal{R} \to \mathcal{Q}, \quad (\rho, \theta) \mapsto \left(\frac{\sin\theta}{\cos\theta}, \frac{(\cos\theta + \sin\theta + \rho\cos\theta\sin\theta)\cos^2\theta}{\sin\theta} \right),$$

satisfies $\psi(\mathcal{R}) \subset \overline{\mathcal{Q}}$ and $g \circ \psi = \phi|_{\mathcal{R}}$. Consequently, $\phi(\mathcal{R}) \subset \mathcal{S}$, as required. $\square$

3.2.2. *The inequality $\phi_1^2(\rho,\theta) + \phi_3^2(\rho,\theta) \geq \frac{\rho^2}{4}$ holds for each $(\rho,\theta) \in \overline{\mathcal{R}}$. Consequently,*

$$\operatorname{dist}(\phi(\rho,\theta), \mathbf{0}) \geq \frac{\rho}{2} \tag{3.7}$$

for each $(\rho,\theta) \in \overline{\mathcal{R}}$.

Proof. As ρ, $\cos\theta$, $\sin\theta$ are ≥ 0 on $\overline{\mathcal{R}}$, we have

$$\begin{aligned} \phi_1(\rho,\theta) &\geq \rho\cos\theta(\cos^4\theta + \sin^4\theta) = \rho\cos\theta(1 - 2\cos^2\theta\sin^2\theta) \\ &= \rho\cos\theta\left(1 - \frac{\sin^2(2\theta)}{2}\right) \geq \frac{\rho}{2}\cos\theta. \end{aligned}$$

In addition, $\phi_3(\rho,\theta) = \rho\sin\theta \geq \frac{\rho}{2}\sin\theta$, so

$$\phi_1^2(\rho,\theta) + \phi_3^2(\rho,\theta) \geq \frac{\rho^2}{4}\cos^2\theta + \frac{\rho^2}{4}\sin^2\theta = \frac{\rho^2}{4},$$

as required. $\square$

3.2.3. The map ϕ satisfies $\phi(0,\theta) = \phi(0, \frac{\pi}{2} - \theta)$ for $\theta \in [0, \frac{\pi}{2}]$. Fix $M > 0$ and consider the rectangle $\overline{\mathcal{R}}_M := [0, M] \times [0, \frac{\pi}{2}]$. Denote $\phi_M := \phi|_{\overline{\mathcal{R}}_M}$. Identify the points $(0,\theta)$ and $(0, \frac{\pi}{2} - \theta)$ for $\theta \in [0, \frac{\pi}{2}]$ and endow the quotient space $\tilde{\overline{\mathcal{R}}}_M$ with the quotient topology. Observe that the interior $\operatorname{Int}(\tilde{\overline{\mathcal{R}}}_M)$ of $\tilde{\overline{\mathcal{R}}}_M$ as a topological manifold with boundary is the quotient space $\tilde{\mathcal{R}}_M$ obtained identifying the points $(0,\theta)$ and $(0, \frac{\pi}{2} - \theta)$ of $\mathcal{R}_M := [0, M) \times (0, \frac{\pi}{2})$, where $\theta \in (0, \frac{\pi}{2})$.

The canonical projection $\pi_M : \overline{\mathcal{R}}_M \to \tilde{\overline{\mathcal{R}}}_M$ is continuous. As ϕ_M is compatible with π_M, there exists a continuous map $\tilde{\phi}_M : \tilde{\mathcal{R}}_M \to \mathbb{R}^3$ such that the following diagram is commutative. In addition, $\tilde{\phi}_M(\tilde{\mathcal{R}}_M) = \phi(\mathcal{R}_M) \subset \mathcal{S}$.

$$\begin{tikzcd} \mathcal{R}_M \arrow[r, hook] \arrow[d, "\pi_M|_{\mathcal{R}_M}"'] & \overline{\mathcal{R}}_M \arrow[d, "\pi_M"'] \arrow[dr, "\phi_M"] & \\ \tilde{\mathcal{R}}_M \arrow[r, hook] & \tilde{\overline{\mathcal{R}}}_M \arrow[r, "\tilde{\phi}_M"'] & \mathbb{R}^3 \end{tikzcd}$$

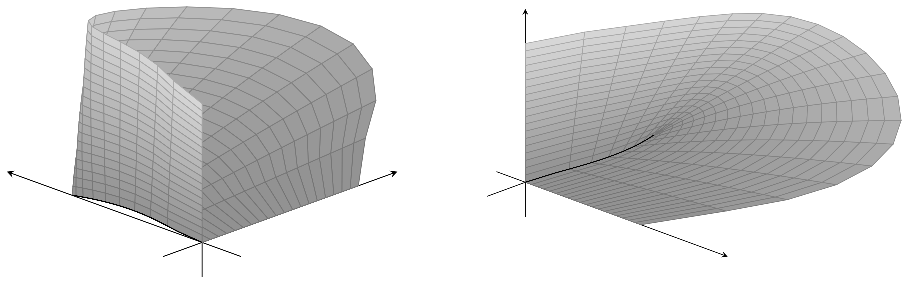


Figure 3: Left and right views of $\phi_M(\mathcal{R}_M) \subset \mathcal{S}$.

3.2.4. *$\tilde{\overline{\mathcal{R}}}_M$ is homeomorphic to a disc and its boundary is the set*

$$\pi_M(\{\rho = M\} \cup \{\theta = 0\} \cup \{\theta = \tfrac{\pi}{2}\}).$$

Proof. Identify $\mathbb{R}^2$ with $\mathbb{C}$ (interchanging the order of the variables $(\rho,\theta) \rightsquigarrow (\theta,\rho)$) and consider the continuous map

$$\mu : \mathbb{C} \to \mathbb{C},\ z := \theta + \sqrt{-1}\rho \mapsto w := u + \sqrt{-1}v = (\tfrac{4}{\pi}z - 1)^2.$$

The restriction $\mu|_{\{\rho>0\}} : \{\rho > 0\} \to \mathbb{C} \setminus ([0,+\infty) \times \{0\})$ is a homeomorphism and the image of $\overline{\mathcal{R}}_M \setminus \{\rho = 0\}$ is

$$\mathcal{T}_M := \big\{(u,v) \in \mathbb{R}^2 :\ (\tfrac{\pi v}{8M})^2 - (\tfrac{4M}{\pi})^2 \le u \le 1 - (\tfrac{v}{2})^2\big\} \setminus ([0,1] \times \{0\}).$$

The closure $\overline{\mathcal{T}}_M$ of $\mathcal{T}_M$ is a compact convex set (as it is a closed bounded intersection of two convex sets). By [1, Cor.11.3.4] $\overline{\mathcal{T}}_M$ is homeomorphic to a closed disc. In addition

$$\mu|_{\{\rho=0\}} : \{\rho = 0\} \to [0,+\infty) \times \{0\},\ \theta \mapsto (\tfrac{4}{\pi}\theta - 1)^2$$

transforms the segment $[0, \frac{\pi}{2}] \times \{0\}$ onto the interval $[0,1]$. The preimage of $t_0 \in [0,1]$ under $\mu|_{\{\rho=0\}}$ is

$$\{\theta_1 := \tfrac{\pi}{4}(1 + \sqrt{t_0}), \theta_2 := \tfrac{\pi}{4}(1 - \sqrt{t_0})\}.$$

As $\theta_1 = \frac{\pi}{2} - \theta_2$, the map $\lambda := \mu|_{\overline{\mathcal{R}}_M} : \overline{\mathcal{R}}_M \to \overline{\mathcal{T}}_M$ factors through $\tilde{\overline{\mathcal{R}}}_M$ and there exists a continuous map $\tilde{\lambda} : \tilde{\overline{\mathcal{R}}}_M \to \overline{\mathcal{T}}_M$ such that the following diagram is commutative.

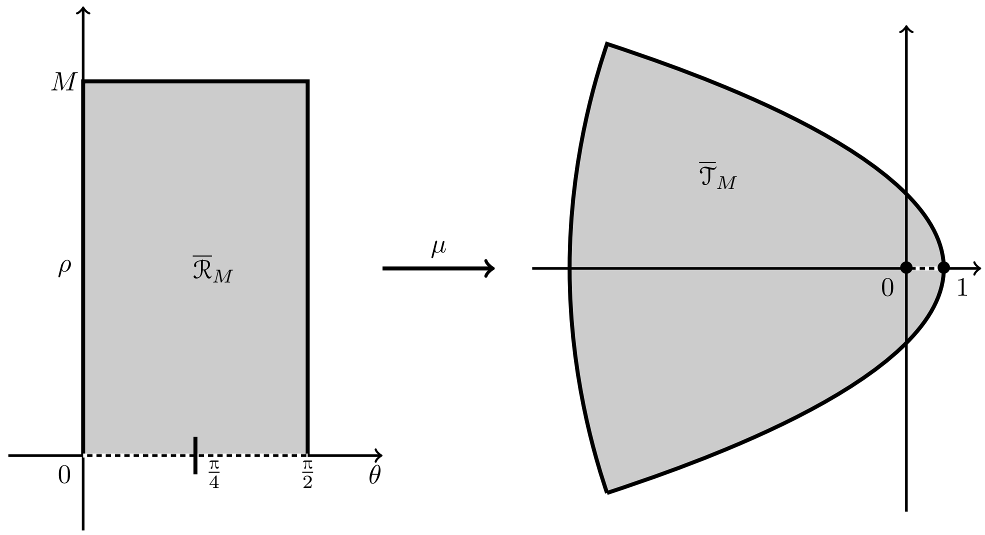


Figure 4: Behavior of the map $\mu : \overline{\mathcal{R}}_M \to \overline{\mathcal{T}}_M$.

$$\begin{tikzcd}
\mathcal{R}_M \arrow[r, hook] \arrow[d, "\pi_M|_{\mathcal{R}_M}"'] & \overline{\mathcal{R}}_M \arrow[d, "\pi_M"'] \arrow[dr, "\lambda"] & \\
\tilde{\mathcal{R}}_M \arrow[r, hook] & \tilde{\overline{\mathcal{R}}}_M \arrow[r, "\tilde{\lambda}"'] & \mathcal{S}_M
\end{tikzcd}$$

The map $\tilde{\lambda}$ is continuous and bijective and it maps the compact set $\tilde{\overline{\mathcal{R}}}_M$ onto the Hausdorff space $\overline{\mathcal{T}}_M$, so it is a homeomorphism. Consequently, $\tilde{\overline{\mathcal{R}}}_M$ is homeomorphic to a disc and its boundary is $\pi_M(\{\rho = M\} \cup \{\theta = 0\} \cup \{\theta = \frac{\pi}{2}\})$, as required. $\square$

3.2.5. *Fix $B \geq A > 0$ and consider the warped discs $\mathcal{D}_1$ and $\mathcal{D}_2$ introduced in* (3.3) *and* (3.4). *Then there exists $M > 0$ such that the boundary map $\partial\tilde{\phi}_M : \partial\tilde{\overline{\mathcal{R}}}_M \to \mathbb{R}^3$ meets transversally once both discs $\mathcal{D}_1$ and $\mathcal{D}_2$.*

Proof. As $\mathcal{D}_1$ and $\mathcal{D}_2$ are bounded set, there exists $M_0 > 0$ such that $\mathcal{D}_1 \cup \mathcal{D}_2 \subset \{\|(x, y, z)\| < M_0\}$. Take $M := 4M_0$ and consider the set $\tilde{\overline{\mathcal{R}}}_M$ and the continuous map $\tilde{\phi}_M$ introduced in paragraph 3.2.3.

We claim: *the boundary map $\partial\tilde{\phi}_M : \partial\tilde{\overline{\mathcal{R}}}_M \to \mathbb{R}^3$ meets transversally once $\mathcal{D}_1$.*

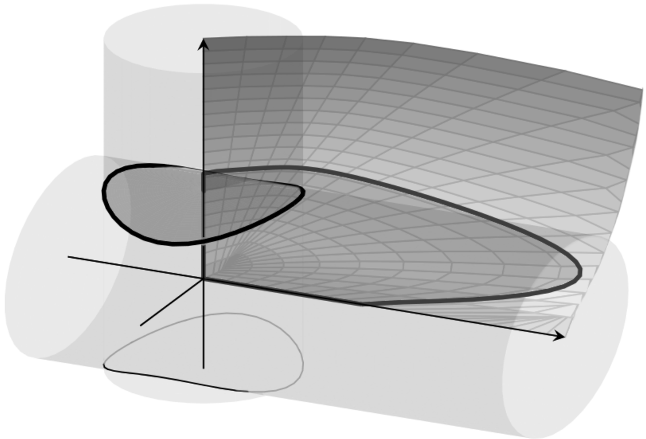


Figure 5: The boundary map $\partial\tilde{\phi}_M : \partial\tilde{\overline{\mathcal{R}}}_M \to \mathbb{R}^3$ meets transversally once $\mathcal{D}_1$.

Consider the parameterization of $\partial\tilde{\overline{\mathcal{R}}}_M$ given by

$$\beta_1(t) := \begin{cases} \pi_M(t, \frac{\pi}{2}), & \text{if } 0 \le t \le M, \\ \pi_M(M, M + \frac{\pi}{2} - t), & \text{if } M < t \le M + \frac{\pi}{2}, \\ \pi_M(2M + \frac{\pi}{2} - t, 0), & \text{if } M + \frac{\pi}{2} < t \le 2M + \frac{\pi}{2}. \end{cases}$$

We have

$$\alpha_1(t) := \tilde{\phi}_M \circ \beta_1(t) = \begin{cases} \phi(t, \frac{\pi}{2}), & \text{if } 0 \le t \le M, \\ \phi(M, M + \frac{\pi}{2} - t), & \text{if } M < t \le M + \frac{\pi}{2}, \\ \phi(2M + \frac{\pi}{2} - t, 0), & \text{if } M + \frac{\pi}{2} < t \le 2M + \frac{\pi}{2}. \end{cases}$$

Choose $0 < \varepsilon < \min\{B, M_0 - B\}$ and consider the homeomorphism

$$\zeta_1 : \mathbb{R}^3 \to \mathbb{R}^3,\ (x, y, z) \mapsto (x, y, z - \xi_1(x, y)),$$

where ξ_1 is the (continuous) semialgebraic function introduced in (3.5). Denote $\mathcal{D}_1(\varepsilon) := \zeta_1^{-1}(\mathbb{D}_A(\varepsilon))$. It is enough to check:

$$\alpha_1^{-1}(\mathcal{D}_1(\varepsilon)) = (B - \varepsilon, B + \varepsilon).$$

Pick $p_0 := \alpha_1(t_0) \in \operatorname{Im}(\alpha_1)$. We distinguish three cases:

(i) If $0 \le t_0 \le M$, then $\zeta_1(p_0) = (\zeta_1 \circ \phi)(t_0, 0) = (0, 0, t_0 - B)$. Consequently, $\zeta_1(p_0) \in \mathbb{D}_A(\varepsilon)$ if and only if $-B < -\varepsilon < t_0 - B < \varepsilon < M - B$.

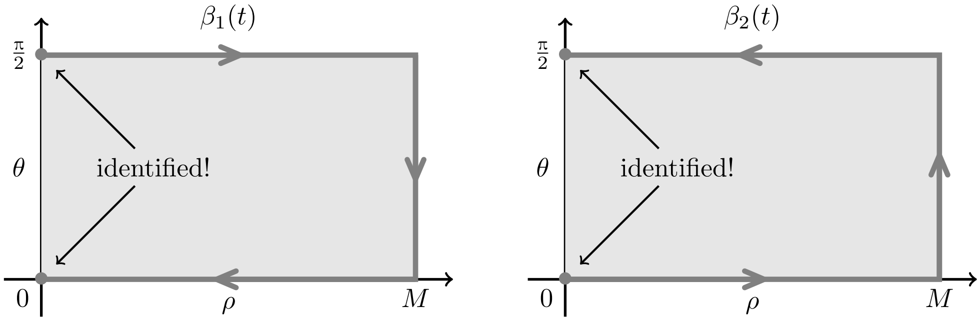


Figure 6: Behavior of the paths β_1 and β_2.

(ii) If $M < t_0 \leq M + \frac{\pi}{2}$, we have by (3.7)

$$\operatorname{dist}(p_0, \mathbf{0}) \geq \tfrac{M}{2} = 2M_0 > \sqrt{2}M_0 > \operatorname{dist}(q, \mathbf{0})$$

for each $q \in \mathcal{D}_1(\varepsilon)$. Therefore $p_0 \notin \mathcal{D}_1(\varepsilon)$.

(iii) If $M + \frac{\pi}{2} < t_0 \leq 2M + \frac{\pi}{2}$, then

$$p_0 = \alpha_1(t_0) = \phi(2M + \tfrac{\pi}{2} - t_0, 0) = (2M + \tfrac{\pi}{2} - t_0, 0, 0),$$

so $\zeta_1(p_0) = (2M + \frac{\pi}{2} - t_0, 0, -B)$. As $\varepsilon < B$, it holds $\zeta_1(p_0) \notin \mathbb{D}_A(\varepsilon)$, so $p_0 \notin \mathcal{D}_1(\varepsilon)$.

We conclude $\alpha_1^{-1}(\mathcal{D}_1(\varepsilon)) = (B - \varepsilon, B + \varepsilon)$, so α_1 meets transversally once $\mathcal{D}_1$.

Analogously one shows: *the boundary map $\partial\tilde{\phi}_M : \partial\tilde{\overline{\mathcal{R}}}_M \to \mathbb{R}^3$ meets transversally once $\mathcal{D}_2$.*

Consider in this case the parameterization of $\partial\tilde{\overline{\mathcal{R}}}_M$ given by

$$\beta_2(t) := \begin{cases} \pi_M(t, 0), & \text{if } 0 \leq t \leq M, \\ \pi_M(M, t - M), & \text{if } M < t \leq M + \frac{\pi}{2}, \\ \pi_M(2M + \frac{\pi}{2} - t, \frac{\pi}{2}), & \text{if } M + \frac{\pi}{2} < t \leq 2M + \frac{\pi}{2}. \end{cases}$$

We have

$$\alpha_2(t) := \tilde{\phi}_M \circ \beta_2(t) = \begin{cases} \phi(t, 0), & \text{if } 0 \leq t \leq M, \\ \phi(M, t - M), & \text{if } M < t \leq M + \frac{\pi}{2}, \\ \phi(2M + \frac{\pi}{2} - t, \frac{\pi}{2}), & \text{if } M + \frac{\pi}{2} < t \leq 2M + \frac{\pi}{2}. \end{cases}$$

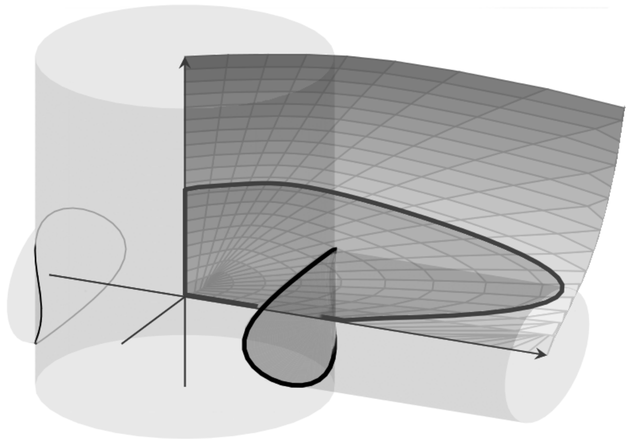


Figure 7: The boundary map $\partial\tilde{\phi}_M : \partial\tilde{\overline{\mathcal{R}}}_M \to \mathbb{R}^3$ meets transversally once $\mathcal{D}_2$.

Proceed as above keeping the same values for A and ε and using in this case the homeomorphism

$$\zeta_2 : \mathbb{R}^3 \to \mathbb{R}^3,\ (x, y, z) \mapsto (z, y, x - \xi_2(z, y)),$$

where ξ_2 is the (continuous) semialgebraic function introduced in (3.6), to prove that α_2 meets transversally once the warped disk $\mathcal{D}_2$. $\square$

3.2.6. By 3.2.4 $\tilde{\overline{\mathcal{R}}}_M$ is homeomorphic to a closed disc. By Proposition 2.5 applied to the continuous map $\tilde{\phi}_M : \tilde{\overline{\mathcal{R}}}_M \to \mathbb{R}^3$ and 3.2.5, we deduce that the boundaries of both warped discs $\mathcal{D}_1$ and $\mathcal{D}_2$ meet $\phi_M(\mathcal{R}_M) \subset \mathcal{S}$. Thus, 3.1.1 holds, as required. $\square$

References

[1] M. Berger: *Geometry, I.* Universitext, Springer-Verlag, Berlin 1987.

[2] J. F. Fernando: *Polynomial images of* $\mathbb{R}^n$*.* In *Reelle algebraische und analytische Geometrie, March 17– 23 (2002)* Oberwolfach.

[3] J. F. Fernando: *On the one-dimensional polynomial and regular images of* $\mathbb{R}^n$*.* J. Pure Appl. Algebra **218,** 9(2014) 1745–1753.

[4] J. F. Fernando, J. M. Gamboa: *Polynomial images of* $\mathbb{R}^n$*.* J. Pure Appl. Algebra **179,** 3 (2003) 241–254.

[5] J. F. Fernando, J. M. Gamboa: *Polynomial and regular images of* $\mathbb{R}^n$*.* Israel J. Math. **153** (2006) 61–92.

[6] J. F. Fernando, J. M. Gamboa, C. Ueno: *On convex polyhedra as regular images of* $\mathbb{R}^n$*.* Proc. London Math. Soc. (3) **103** (2011) 847–878.

[7] J. F. Fernando, J. M. Gamboa, C. Ueno: *Sobre las propiedades de la frontera exterior de las imágenes polinómicas y regulares de* $\mathbb{R}^n$*.* In *Contribuciones Matemáticas en homenaje a Juan Tarrés,* 159–178, UCM 2012.

[8] J. F. Fernando, C. Ueno: *On complements of convex polyhedra as polynomial and regular images of* $\mathbb{R}^n$*.* Int. Math. Res. Not. **18** (2014) 5084–5123.

[9] J. F. Fernando, C. Ueno: *On the complements of* 3*-dimensional convex polyhedra as polynomial images of* $\mathbb{R}^3$*.* Internat. J. Math. **25,** 7 (2014) 1450071, 18pp.

[10] J. F. Fernando, C. Ueno: *On the set of points at infinity of a polynomial image of* $\mathbb{R}^n$*.* Discrete Comput. Geom. **52,** 4 (2014) 583–611.

[11] J. F. Fernando, C. Ueno: *A short proof for the open quadrant problem.* Preprint RAAG (2014, submitted to MEGA 2015) 8 pp.

[12] J. M. Gamboa: *Algebraic images of the real plane.* In *Reelle algebraische Geometrie, June 10-16 (1990)* Oberwolfach.

[13] A. Harnack: *Über die Vielheiligkeit der ebenen algebraischen Kurven.* Math. Ann. **10** (1876) 189–198.

[14] S. Pinchuk: *A counterexample to the real Jacobian Conjecture.* Math. Z. **217** (1994) 1–4.

[15] A. Tarski: *A decision method for elementary algebra and geometry.* Prepared for publication by J.C.C. Mac Kinsey, Berkeley 1951.

[16] C. Ueno: *Imágenes polinómicas y regulares de espacios euclídeos.* Ph.D. Thesis UCM 2012.

[17] C. Ueno: *A note on boundaries of open polynomial images of* $\mathbb{R}^2$*.* Rev. Mat. Iberoam. **24,** 3 (2008) 981–988.

[18] C. Ueno: *On convex polygons and their complements as images of regular and polynomial maps of* $\mathbb{R}^2$*.* J. Pure Appl. Algebra **216,** 11 (2012) 2436–2448.

A la búsqueda de la espiritualidad perdida

Meditar itinerante acerca del número, el tacto, la duración y el Arte so pretexto de las Matemáticas y la matematización del mundo

Jesús FORTEA PÉREZ

Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
28040 Madrid, España
jfortea@ucm.es

Al Profesor José María Montesinos Amilibia, con motivo de su jubilación.

Resumen

El artículo presente viene motivado por el tema del tacto en su relación con las Matemáticas, si bien se inicia refiriéndose a la palabra y al número, para luego pasar a ocuparse del tacto y la extensión, prosiguiendo con la "durée" bergsoniana y otros desarrollos conexos. El término "espiritualidad", que figura en su título, ha de entenderse en el sentido más amplio posible, abarcando, en particular, cuanto la tecnología ha hecho perder al hombre, extremo que el artículo crítica desde varias perspectivas. Su entender del término espiritualidad es, pues, acorde con lo enseñado por el "segundo Heidegger", si es que hemos de seguir a Derrida. Los Post Scripta que lo concluyen buscan expresarse más con imágenes que con razonamientos y tratan, en particular, del Arte-Moderno.

1. El abandono de la palabra

(1.1) De las dos gnosis supremas de salvación que, en halo de modernidad, se ofrecían al pensar universitario español cuando el Profesor Montesinos Amilibia inició su dedicación docente, no otras sino el marxismo y la que, en sus diversas variantes, giraba en torno a la ciencia positiva como paradigma, es ésta la que en la actualidad persevera, habiendo incluso acentuando su carácter de gnosis salvadora. Fundamentalmente, porque la ciencia positiva constituye el soporte de la tecnología moderna de la que hoy en día la sociedad en su conjunto lo espera todo. Una tecnología que, a diferencia de la no matemática propia del mundo antiguo [1], ha alcanzado su éxito, dado el pronto asumir por la ciencia moderna del parecer de Galileo, expuesto en "Il Saggiatore" y según el cual «*La filosofia è scritta in questo grandissimo libro... Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi, ed altre figure geometriche, senza i quali mezi è impossibile a intenderne umanamente parola*» [2]. Ciencia y Tecnología modernas se mostraron, pues y desde muy temprano ajenas a la enseñanza de Lutero el cual reconocía en la palabra que no en el número y la geometría a la fuente suprema de la verdad sobre la Naturaleza y en Adán al más grande de los filósofos por capaz de nombrar, en uso de la palabra, a los animales tras la Creación [3]. Y puesto que Leibniz, en absoluta coherencia con esa enseñanza, consideraba que cuanto más próxima se encontrase una lengua de la de Adán, tanto mayor habría de ser su correspondencia con la verdad de la Naturaleza, se comprende que, para él, encontrarla o reconstruirla resultase la vía científica que seguir y no la propuesta por Galileo. Y es que, según Lutero, la especie humana después de que en el Pentecostés recibieran los Apóstoles el "don de lenguas", había recuperado el estado de Adán perdido tras la confusión de Babel [4]. Nada tiene de sorprendente, entonces, que la ciencia moderna se distanciase también de Leibniz, cuya noción de mónada, sabiamente enraizada en la de forma substancial aristotélica, hacía de él un metafísico, circunstancia que, en sus análisis, parece marginar Bertrand Russell sapientísimo incrédulo famoso. Todo ello sin menoscabo de que Leibniz es asimismo y con Newton el creador del cálculo infinitesimal, si bien —por partir de Nicolas de Cusa y Kepler—, de modo diferente, como lo recuerda el Non-Standard Analysis de Robinson [5] con sus definiciones de infinitésimo y de mónada.

(1.2) Abandonado, pues, todo género de trato con la palabra, incluido el que conllevaba el conocimiento de la Cábala —tan cara a Leibniz, a Lady Conway o a More dada la lógica proximidad que atribuían al hebreo respecto de la lengua de Adán [6]—, la Naturaleza llegó a ser considerada en las Islas británicas como el *"unsealed book of God"*, si bien significativamente escrito en clave matemática que no verbal. Este carácter matemático poco tenía que ver, sin embargo, con las ideas innatas de sesgo platónico invocadas por Descartes, ya que el *"hipotheses non fingo"*, de Newton [7] —criticado luego por Heidegger— en su condición de manifestación señera de lo que no era sino una *"Revolt against Rationalism"* [8], lo que buscaba era liberarse de la metafísica del siglo XVII y, en particular, de la cartesiana. Todo ello, además, cuidando muy mucho la vertiente aplicada del conocimiento. De ahí que Boyle deseara que la Royal Society triunfase en su intento por *"discover the true nature of the works of God"*, de modo que sus logros y los de otros investigadores pudieran ser referidos no ya *"to the glory of the great Author of Nature"* sino también y muy principalmente *"to the confort of mankind"* [9].

(1.3) Había surgido, así, en apreciación de Richard Bentley —canónigo de humanismo cierto por versado en lenguas clásicas— una *"filosofía mecánica"* [10]. Muy alabada en el continente por Madame de Chatelet y Voltaire frente al cartesianismo metafísico [11], pronto suscitaría, sin embargo, las dudas del mismo Bentley y de muchos otros. Todos ellos recelando de que, a diferencia de la filosofía que, siguiendo a

Lutero, pretendía reconstruir la lengua de Adam, resultase bastante más alejada del *"great Author of Nature"* de lo que proclamaba. Y es que, para esta filosofía y en la línea de lo apreciado por Bentley, el mundo no era ya sino una máquina poco menos que automática. Atendida todo lo más por un *"divino relojero"*, su existencia resultó cada vez más problemática, al punto de llegar a considerarla Laplace —cuentan que en respuesta a preguntas de Napoleón—, como una hipótesis enteramente prescindible. No en vano poco antes la diosa Razón había sido entronizada en Notre Dame. De nada serviría, por lo demás y más tarde, que Goethe tratase de reparar el ultraje que Newton había perpetrado contra la luz destruyéndola, sin más, con un prisma para, en ese estado, estudiarla [12]. Por supuesto tampoco que, incluso antes, Schiller invocase en alabanza a los dioses griegos contraponiéndolos al Cristianismo [13] o que Novalis —él sí básicamente cristiano— escribiera en sus *"Hymnen an the Nacht"* que «Donde no hay dioses, reinan los fantasmas» («*Wo keine Götter sind, walten Gespenster*») [14]. De hecho, la hipótesis del *"divino relojero"* era, para Novalis, una forma de desalojar a la Gracia del mundo, trocándola por la Tecnología y sus encantos de Falsirena gracianesca. De este manera y en su criterio, en un mundo *desencantado* con toda huella de lo sagrado eliminada, «*el hombre moderno podría dedicarse sin descanso a limpiar a la Naturaleza de toda traza de poesía*» [15]. Y es que el *monoteismo de la razón* —ese monoteísmo del que ya se ocuparon Reinhold y Schiller considerándolo transmitido por los sacerdotes egipcios a Moisés para que, por su conducto y en forma de monoteismo hebreo, pasase al Cristianismo asegurándole así una difusión universal—, había, vía la diosa Razón, alcanzado plenamente su objetivo, no otro sino el ateísmo [16]. Al menos su coincidencia con el *"Deus sive Natura"* del cartesiano metafísico Spinoza permitía calificarlo de este modo [16].

2. Número y harmonía

(2.1) Semejante implicación de las Matemáticas en la desacralización del mundo como la descrita, pone de manifiesto un cambio cualitativo surgido en las mismas. Pues fue justo la inmanencia del número, en invocación explícita de las deidades del Olimpo, lo que había caracterizado al Pitagorismo en sus diversos desarrollos e influencias. Sin ir más lejos, en la influencia que ejerció sobre Platón, cuyo Timeo contemplaba un Cosmos *vivo y divino* [17], resuelto en esferas bellas, el magno Cielo estrellado su piel tersa y profunda [18], toda ella de catasterismos irisada, proclamando las cuitas y amores entre humanos, ninfas y deidades [19]. O lo que es lo mismo, dando testimonio de la presencia de la Transcendencia en la Tierra. Concepción del Cosmos que sostuvo la piedad pagana de Proclo neoplatónico y matemático, estando el Mundo clásico ya en su crepúsculo, la Escuela de Atenas a punto de ser cerrada y Simplicio yendo a morar en la Harrán de los dinteles en caracteres siriacos gloriando las enseñanzas del Timeo [20]. Tenuamente presente en tiempos de Carlomagno gracias a Calcidio neoplatónico sucesor de Osio obispo en Córdoba, brotaría con ímpetu nuevo en Fiésole y en alabanza de Proclo [21]. Y así su elan renacido llegó a Praga la de

Rodolfo II, el Cosmos de nuevo en ella conociendo una nueva resolución geométrica y musical pitagórica, ahora por obra de Kepler. Kepler el hacedor de horóscopos, se supone que para disgusto de Adorno, rancio intelectual melómano moderno [22], de madre católica ufanándose de haber cantado en la Ópera Imperial de Viena. Enemigo de Heidegger [23], todo parece cuadrar, ya que significativamente Heidegger no era especialmente querencioso de la música y sí, en cambio, de la poesía [24].

(2.2) Dado que los matemáticos no son genios sino talentos [25], si se llegase a pensar que el descrito rechazo de la concepción del Cosmos que Pitágoras y Platón legaron incurre en sectarismo y parcialidad, sería, no obstante, posible acordar un compromiso con los matemáticos. En efecto, bastaría con ver en la nueva concepción del Cosmos propugnada por ellos, no un genial entendimiento —ni que decir tiene que sectario—, sino un *mero alarde talentoso.* Es más, el tal alarde habría dado paso —con la inmanencia del número por completo perdida—, a ese *talentudo* concierto instrumental *sin voz ni palabra alguna* —los músicos son asimismo talentos [26]—, en el que las Matemáticas modernas —para no pocos formando parte de la Fisica moderna [27]—, han convertido al mundo. Concierto instrumental, por supuesto, *sin voz ni palabra alguna* a cargo de «*talentos de chicharra en salas de concierto para instrumentos de cuerda*» [28]. Justo, por ello, concierto del gusto de Galileo Galilei, pero opuesto al gusto de Galileo padre, fundador con Monteverdi de la Camerata Fiorentina y coincidente, por tanto, con el de Aristoxeno renegando del pitagorismo por acusarle de tener *a la música como cosa del número que no de la voz y el sentimiento* [29]. Y es que no es misma música la de la *palabra* que la del instrumento, por sujeto éste más al número que la voz.

(2.3) Concierto el de las Matemáticas modernas, como poco dodecafónico y por “superación” de las leyes de la armonía. Pues careciendo ya las funciones de nombres y definiciones por haber sido degradadas a meros puntos de “infinitas” coordenadas, ese carecer determina que no viniesen ya definidas por un criterio, habiéndose, pues, de esperar a que fallezcan para asignarles una lápida definitoria, cosa que nunca sucede, ya que el continuo de los matemáticos formalistas —la inmensísima mayoría— lo tiene así sentenciado. Como consecuencia, su valor en un punto dista de determinarlas, con lo que las tales funciones —por no estar bautizadas—, resultan de lo más lejano a una melodía, melodía en la que, según Bergson, el conocimiento de una nota hace que surja al momento ella toda [30]. Justo es reconocer, no obstante, que los matemáticos, tras arduos trabajos, consiguen mediante difeomorfismos, homeomorfismos y otras industrias, remitir sus desarrollos a funciones y configuraciones en invocación pitagórica genérica de esas *curvas* que son las hipérbolas, elipses o parábolas [31]. Nunca se insistirá bastante que, en Grecia no todo garabato tenía rango de curva, distinguiéndose incluso entre curvas propiamente dichas y curvas *mecánicas* . Sí, por cierto que eran curvas las hipópedes y los óvalos, en tanto que secciones del toro, como las cónicas lo eran del cono con independencia de otras caracterizaciones paralelas. Y fue significativamente una hipópede sobre una esfera que generaba la rotación de otras dos esferas concéntricas de igual radio que la primera girando cada una en sentido contrario sobre dos ejes cruzándose en ese centro, de la que se valió Eudoxo en su intento por enriquecer ese Cosmos, resuelto en esferas, presente tanto en Platón como en Aristóteles. Los tiempos de Euler *talentudo, que no genial* con su moderno concepto de función, estaban aún muy lejos, se dirá que para desgracia de la Modernidad; ella presunta salvadora de la Humanidad, supuesto que semejante vocablo signifique algo.

3. Tacto, topología y escultura

(3.1) «*Il est de parfums frais comme des chairs des enfants*» escribió Baudelaire [32], pero también voces que acarician. Surgidas de la carne en emoción, son voces que apelan, pues, a la nada pitagórica música de Aristoxeno, pero también y esencialmente al tacto y, por tanto, a la mano humana. Ese tacto del que la voz así se acuerda, sabe de la extensión y se ha afirmado, en poco preciso uso de la correspondiente noción cartesiana, que la *"res extensa"* no es sino un *"girón de sensación"* [33]. Por otra parte, S. H. Rosen [34] sostenía, refiriéndose al tacto en su relación con el pensar, que Aristóteles compara en De Anima (432a1) a la *psyché noética* con la *mano humana*, viendo en la mente una mano mental que permite asir el *"ειδοσ"* en la cosa. Por su parte, I. Mueller [35] entendía que, según Aristóteles, los objetos matemáticos no son reales como las substancias sensibles, pero que están conectados con lo sensible, correspondiendo a la abstracción alcanzar desde él la extensión pura, la cual no es sensible, pero no tan indiferenciada como la materia prima tridimensional a la que se refirió Moderato de Gades y luego Simplicio y Filopono [36].

(3.2) De estar Aristóteles en lo cierto, la sensación haría, pues, acto de presencia en el corpus de las Matemáticas, al punto de que, por ello, I. Mueller [37] apreció falta de rigor en Aristóteles por seguir considerando silogístico puro al razonar matemático. Pero evidentemente no sólo como imagen visual propia de geometrías proyectivas divulgadas en decorados renacientes, sino como continuo; esto es, como nacida del tacto en el ámbito tridimensional donde se desarrolla el operar háptico propio de las manos [38] del que, de ningún modo, es ajeno el Analisis Situs de Poincaré [39]) proponiendo a la topología que con él nacía, los problemas que plantea la modelación de las formas. La topología sería, pues, háptica y la "intuición del abstracto", que llevó a Dieudonné [40] a escribir deliberadamente un libro de análisis matemático sin figura alguna, quizás tributaría de ese tacto que en las yemas de los dedos mora [41]. Las Matemáticas entroncarían, entonces, con la escultura. Y es que Herder, siguiendo el camino abierto por Berkeley para el tacto en detrimento de la visión, vino en afirmar que «*todo lo referente a la belleza de una escultura no es concepto visual sino tangible*» [42]. De hecho, las sensaciones táctiles se funden en la pintura antigua con las hápticas e incluso con las instrumentales, tal sucede con los pintores y al tiempo tallistas españoles de los siglos XVI y XVII [43].

(3.3) Todo esto lo admitirían tanto Riegl [44] como Berenson, este último en términos de *"sensaciones ideadas"* [45]. Mas pronto el embeleso se esfuma. Pues es voluntad de los matemáticos que no son intuicionistas, que al continuo se le ha de tener por constituido por puntos inextensos en acto. Ahora bien, si así es, el modelar del continuo no puede llegar a ellos, ya que estos inextensos puntos surgen actualizando su potencialidad —Aristóteles preferiría expresarse así— cuando, no los dedos de la mano sino el instrumento, corta o pulsa el continuo lineal en cortaduras de Dedekind [46]. Son, pues, puntos sin extensión que se escapan de los dedos como el agua de las manos sin que el modelar háptico del continuo pueda retenerlos. Ya de hecho Protágoras advirtió que el que la tangente tocase a la circunferencia —lo mismo cabe decir de la esfera y el plano tangente— en un único punto era cosa que nadie había percibido, de donde deducía que era imposible [47]. Y es que pretender tocar un punto sin extensión conlleva un tocar que no toca en extensión, lo que atenta evidentemente a la noción del tocar. Pues, de la misma forma que Husserl ya advirtiese, haciendo suya la teoría de Stumpf sobre la Verschmelzung, que no hay color sin extensión ni extensión sin color [48] y como pensaban los pitagóricos [49], cabría afirmar parafraseándolo que tampoco hay movientes yemas sin extensión ni extensión que a su caricia no evoque. Con todo, si la cuestión se enfocara desde el pensar del primer Bergson, el que el espacio tenga algo que ver con la sensación, no es admisible [50]. Ahora bien, también es cierto que Bergson, en algunos textos posteriores acepta que sería preciso distinguir entre percepción de la extensión y concepción de espacio, lo que obliga a su vez a distinguir entre espacio vivido y espacio abstracto evitando, de este modo, una separación radical afectando al tiempo y la extensión [51]. De hecho, la percepción se muestra rebelde a dicha separación [52]. *La sensación apunta, pues, a la durée.*

4. Sensualismo, lenguaje y lenguaje matemático en el siglo XVIII europeo

(4.1) Protágoras, expresándose del modo indicado, no hacía sino formular un juicio acorde con la corriente sofista a la que pertenece y emparentada con Isócrates, corriente llamada a revivir luego, en torno a la palabra que no al número, en el humanismo del Renacimiento, humanismo asimismo complacido en la doxa, la sensación y el tacto. Pero era también un sensualismo elegante aquel que, demorado en porcelanas y sedas ornando con galanura mesas y consolas en los palacios de Europa, había acogido esa *"Revolt against Rationalism"* que Newton impulsara con su *"hipotheses non fingo''* y que a Bentley tanto había inquietado. Sensualismo elegante, pues, el de Condillac e inspirando en Rousseau su nunca escrita *"morale sensitive"*, eco temprano que habría sido de Héloïse en lozanía [53]. Y fue Condillac el que, para satisfacción de matemáticos, manifestó que sin signos convencionales previos en el lenguaje no hay posibilidad de que el pensamiento brote desde él y se desarrolle. Ya más dudoso resulta que los matemáticos aceptasen verse representados en la estatua —el prisma newtoniano separando las sensaciones de Condillac—, cuando, dejando ella de serlo, se valía, no ya de la vista sino esencialmente del tacto, en el explorar háptico local mediante sistemas de entornos de su derredor continuo [54]; es decir, similarmente a ese modo topológico de comportamiento o de exploración “local” del espacio que Piaget descubrió mucho más tarde en los niños pequeños y que relacionó con las estructuras madres de Bourbaki [55]. Y no debe causar sorpresa que así fuere, ya que cumplió a Condillac ser en Parma el preceptor de un todavía niño, no otro sino el Infante Fernando nieto de Felipe V de España y de Isabel Farnesio [56].

(4.2) Como Condillac expresamente reconoció, su concepción del lenguaje enlaza con la Grammaire y la Logique de Port Royal. Sería, sin embargo, erróneo deducir de ello que lo hace con el cartesianismo metafísico —Leibniz dixit— de Antonie Arnauld, coautor de dichas obras [57] y a pesar de lo supuesto por Chomsky al hablar de lingüistica cartesiana en referencia a Port Royal. Y es que tanto la Logique como la Grammaire de Port Royal tributarían obviamente y compartiéndolo con Descartes, del agustinismo en general, pero esencialmente del jansenismo, incluyendo por ello la Logique toda la temática de Pascal acerca del infinito en sus dos modalidades, aparte de consideraciones sobre lo efímero del pensar matemático [58]. De hecho la querencia de Condillac por la Gramática de Port Royal tendría significativamente mucho más que ver con Locke estudioso admirador de Port Royal, cuyas obras no sólo leía sino que poseía [59]. No despreciando Condillac ni a la Grammaire ni a la Logique de Port Royal sino todo lo contrario, discreparía, sin embargo, de ellas porque se decataban más por el juicio que por el razonamiento [58], de suerte que, viendo Arnauld y Nicole en la Gramática el arte del discurso y en la lógica el del pensamiento, Condillac optó por ver ya en la Gramática la primera parte del arte de pensar [60] con lo que ello suponía de “algoritmización” del pensamiento. De hecho, Condillac, en su posterior *“Langue des calculs”*, afirmará que «*l’Algebre est une langue bien fait, et c’est la seule: rien n’y paroît arbitraire*» y que «*une science bien traitée n’est qu’une langue bien traitée*» proponiéndose en dicha obra mostrar «*comment on peut donner à toutes les sciences cette exactitude qu’on croit être le partage exclusivif des mathématiques*» [60].

(4.3) La psicología de Condillac, a diferencia de la de Diderot con la que se la ha querido emparentar, no es materialista e implica una espiritualidad clara [61]. Si todo es sensación —los extremos se tocan— todo puede ser espíritual. Berkeley cuentan que se lo hizo ver a Malebranche metafísico, mostrándole que su ocasionalismo quedaba superado con tal de olvidar la *res extensa* cartesiana, cuya existencia, además, no imponía la Biblia [61]. Diderot, por su parte, haría lo mismo con Condillac, teniéndolo, como a Berkeley, por idealista, dado que es imposible demostrar la existencia de los objetos exteriores si todo se reduce a sensaciones del sujeto [62], un problema, en cualquier caso, que, Descartes teniéndolo por fácil, parecía no desconocer al afirmar que «*ya no falta más que examinar si existen las cosas materiales*» [61]. Sea como fuere, si al discurso de la sensación en el hombre se le priva de la instrumentalidad propia del homo faber, puede devenir espiritual, siquiera sea porque primates y ser humano comparten genéricamente la inteligencia práctica instrumental, aunque no el habla. Sucede, además, que la inteligencia instrumental propia del homo faber no sólo se expresa en útiles e ingenios de naturaleza material, sino que también participa en la elaboración de creaciones no materiales como, por ejemplo, la lógica aristotélica, cuyo carácter instrumental ya fue advertido por Gentile al afirmar que el Organum era un «*metodo-strumento,... prima di affrontare il cimento della cognizione*» [63].

5. Homo faber y Matemáticas

(5.1) Si pues hasta el Organum de Aristóteles no es ajeno a la inteligencia instrumental, cabe preguntarse sobre lo que significa hacer del pensar un construir instrumental sustentado en cimientos. Mas, ¿no es eso justo lo que le sucede con las Matemáticas una vez que se constituyeran, vía dialéctica platónica, en un edificio cimentado en axiomas, hipótesis y hasta movedizos postulados por más que Platón proclamase que es el conocer y no el construir lo propio de la geometría? [64]. Y no sólo se está ante una realidad instrumental de naturaleza arquitectónica, sino ingenieril, como pone de manifiesto el triturar del espacio, por parte de los matemáticos, para fabricar luego, con la pulpa resultante, tableros de conglomerado que llaman planos. Es más, esos matemáticos ya se denominan "ingenieros matemáticos", prosperando en los mercados como expertos en "modelización", incluida la que facilita el ordenador, cuyas teclas "digitales" no se tocan sino que se golpean, como si la mano humana fuese «*artejo de insecto que no solar de las yemas en el palpar de los dedos*» [28]. Ya de hecho Scheler comentaba que la diferencia entre Edinson como científico y un chimpancé listo era meramente de grado que no cualitativa [65]. Y es que los monos de Köhler —no desde luego el de Kafka que hasta llegó a académico [66]—, piensan instrumentalmente para ganarse la fruta de cada día, aunque significativamente no por ello hablen. Kantianamente expresándose y de acuerdo con una tradición lingüistica de la que Révész —decidido defensor de la especificidad humana del habla— participa, la descrita instrumentalidad pertenecería a *Verstand* y la palabra, palabra que los monos no alcanzan, a la *Vernunft* [65].

6. A propósito de Maine de Biran, Vico, Ravaisson y Bergson

(6.1) Maine de Biran, partiendo del sensualismo espiritual de Condillac y por vía de introspección profunda en la vivencia del esfuerzo corporal en su relación con el principio de causalidad, vino en descubrir, allá en sus heredades de Gratum Lupis a la vera de Bergerac [68], que, en el fondo de su conciencia, existía un mundo íntimo de pensamientos y querencias que el mundo de las sensaciones no revelaba de ningún modo. Maine de Biran hubo, pues, de distanciarse de Condillac, imputándole el «construir la ciencia con elementos artificiales o lógicos, como se construyen las fórmulas algébricas» [69]. Esta crítica, apuntando a la concepción algébrica y gramatical del lenguaje por parte de Condillac y esencialmente ligada a la función que en ella se concede a los signos para el emerger del pensamiento, es por completo acorde con la concepción de la "inteligencia" en Bergson, cuyo nacimiento él lo sitúa en el reemplazar de las cosas por signos artificiales en ámbito de tiempo por completo espacializado al margen por completo de la *"durée"* a la que sólo por vía de introspección profunda, permite acceder la intuición. Coadyuvó, de este modo, Maine de Biran al emerger de Bergson. Pero también de Ravaisson maestro de Bergson, proclamando en escritos y dibujos esa "$\epsilon\upsilon\rho\upsilon\theta\mu\acute{\iota}\alpha$" [70] que a toda grácil bailarina asiste en contraste con el "$\pi\alpha\theta\acute{o}\varsigma$" maquinal excitado de las marionetas de Kleist; Kleist, el hechizado por Pentesilea amazona, allá en las brumas de Germania, los góticos carrillones escuchando [71]. Y gracia, en suma, sobre la que en sonrisa Mona Lisa le reveló a Leonardo: «*Observa en el serpentear de toda cosa —si es que la quieres conocer y representar bien—, la especie de gracia que le es propia*» [72]. Pues no es la gracia atributo de las máquinas sino realidad que invoca al *"effort générateur de la nature"*. Y así lo reconoció Bergson, teniendo al retrato de Mona Lisa como debido, no a la artesanía mecánica, sino al Arte [73].

(6.2) Para Bergson, expresándose en palabras de Gouhier, «*si l'esprit ne se sent pas inmédiatement dans sa spiritualité, c'est précisément parce que les exigences de l'action l'obligent à penser les choses avant de se penser lui-même, de sorte qu'il traduit la qualité en quantité, le devenir en immovilité, le fluide en solide, la durée en espace*» [74]. Parece, pues, que, desde una perspectiva bergsoniana —en concreto la del Bergson primero— las Matemáticas no constituirían precisamente un camino de espiritualidad. Y es que en ellas se manifiestan esas exigencias de la acción, en términos de operatividad y movimiento propias del homo faber, a las que tanto su relación con la Física como su razonar propio, le obligan, tal ilustra, sin ir más lejos, la Geometría con su fundamentación arquitectónica. Pero no sólo en esa fundamentación aisladamente considerada, ya que, aparte de la existencia de grupos de transformaciones e invariantes al modo Erlangen, sucede lo mismo con la Teoría de Números. Pues aunque, para Kant, los números tributan de la temporalidad como Brouwer se haría más tarde eco, el razonar de la Teoría de Números resulta inconcebible sin la Topología y la Geometría, las dos a su vez inconcebibles sin referencia al movimiento y al espacio. Ya significativamente Vico retórico aislado y genial advertía en su *Scienza Nova* que siempre propendemos a pensar en términos de corporeidad [75]. Vico, expresándose de ese modo, se mostraba en coincidencia con los pitagóricos, ya que éstos representaban espacialmente los números. Una tradición pitagórica que emerge de nuevo en Bergson, para el que «*toute idée claire de nombre implique une vision dans l'espace*» [58], siendo, por lo demás, completo el acuerdo entre pitagóricos y Bergson en cuanto a la representación del número dos y el *"déroulement"* de toda multiplicidad numérica en el espacio, pero sin olvidar que, para los pitagóricos, las unidades son corpóreas [76].

7. Bergson, Einstein, de Broglie y el problema del espacio y del tiempo

(7.1) «*La science positive consiste essentiellement dans l'élimination de la durée*» afirmaba Bergson, rememorando en carta remitida a Williams James el 9–V–1908 la finalización de sus estudios en l'École normale [74]. Por aquel entonces, Bergson se inclinaba por negar incluso la realidad del tiempo en el mundo material, si bien ya en *Matière et Memoire* se refería explícitamente a *"la durée des choses"*. Como consecuencia, el mundo material se configuraba, para él, como una *"durée diluée"* o *"étendue"*, de suerte que «*l'espace statique n'est qu'une limite idéale de la durée d'étendue*» [77]. El mundo extenso de la materia se mostraría, así, en tanto que sucesión temporal y las denominadas partículas materiales no serían sino cortes arbitrarios en la totalidad del devenir físico [78], careciendo, en consecuencia, de sentido que la "inteligencia" se plantee el problema de su localización precisa y en consonancia con lo afirmado por esa "ciencia positiva" que era la Mecánica Cuántica [79] antes de que de Broglie terminase por asumir las tesis deterministas tras su adhesión a las tesis de Bohm [80]. Desde el punto de vista relativista y aun cuando Bergson no se percatara de ello debido a su presunto mal entender de las tesis de Einstein sobre la simultaneidad, incluso el espacio cuatridimensional de Minkowski donde la "inteligencia" juega a favor de esa *"étendue"* en tanto que *"durée diluée"*. Y es que muestra cómo la "inteligencia" hace del universo una realidad en la que el tiempo —aunque espacializado y descompuesto en instantes sin duración y, por tanto, inadmisibles para Bergson—, aparece indisolublemente amalgamado con el espacio, mostrando, en particular, como carente de sentido todo corte instantáneo tridimensional del universo, resultando, de este modo, imposible la existencia, en feliz afirmación de Eddington, de instantes tan vastos como el mundo [81]. Sería Whitehead el llamado a proponer, desde su profundo conocimiento de las Matemáticas, una concepción filosófica del Universo grandemente influida por el pensamiento de Bergson [82].

(7.2) *"Durée diluée"* en un *"universe qui dure"* cuyo conocimiento es posible gracias a la intuición entendida como «*sympathie par laquelle on se transporte à l'intérieur d'un objet pour coincider avec ce qu'il a d'unique et par conséquent d'inexprimable*» [83]. Por tanto, conocimiento muy distinto al de la "inteligencia", ya que el análisis en el que ella se fundamenta «*ramène l'objet à des éléments déjà connus, c'est à dire communs à cet objet et d'autres*» con lo que se expresa «*une chose en fonction de ce qui n'est pas elle*» [83]. Poco parece tener, pues, que ver la *"intuition"* con ese homo faber tan emparentado con el matemático que, según el parecer de Thom matemático [84], destruye previamente lo que quiere conocer, levantando luego con los despojos y osamentas de su carnicería, un edificio a modo de máquina o artefacto que pretende ser la réplica mimética de lo que quería conocer y que ahora se piensa que, por ser obra suya, ya la puede conocer. Obra, por tanto, de la inteligencia ignorando ese *"elan vital"* que fusiona inteligencia e instinto y sin que el Arte sustituya en ella al *"effort générateur de la nature"* [85], resultando al respecto de lo más revelador que sea Escher, con sus composiciones geométricas hiperbólicas —todo lo más a la altura de las abstractas azulejerías islámicas [86]—, el único artista denominado del que se acuerdan los matemáticos a la hora de generar confusión a cuenta de lo que es el Arte. (Véanse Post scripta al final del escrito.)

8. Pensamiento personal y Matemáticas

(8.1) Sería, sin embargo, equivocado no apreciar en esa *"durée étenduée"* en un *"universe qui dure"*, la posibilidad de una ciencia y filosofía de la Naturaleza. Llevarla a término fue quizás lo que motivó a Bergson para escribir *"L'Évolution créatrice"*. Es más, para Bergson —en texto escogido por Barthélemy-Madaule— «*la matière et la vie qui remplissent le monde sont aussi bien en nous; les forces qui travaillent en toutes choses, nous les sentont en nous; quelle que soit l'essence intime de ce qui est et de ce qui se fait, nous en sommes*» [87]. En otros términos: esa ciencia y filosofía de la Naturaleza pretendida ha de estar necesariamente incardinada en una dinámica psicológica personal como sucedía, por ejemplo, con la Alquimia, invitando la "intuición temporal" de Brouwer a verla quizás presente en las Matemáticas entendidas al modo neoplatónico vivencial. Así, se ha querido ver, en cualquier caso, en referencia a las ciencias de la Naturaleza con Teilhard de Chardin [86]. Mas por mucho que el recuerdo de la cosmoliturgia neoplatónica en Máximo el Confesor venga entonces a la memoria [88], no por ello ha de dejar de constatarse que, de hecho y a pesar de los esfuerzos de René Thom entre otros, cuanto más se matematiza la ciencia presente, más se aliena en el sentido marxista del término. Réplica, por tanto, de la Naturaleza, ha de entrar, entonces, esta ciencia matematizada y a modo de doble, en mimesis conflictiva con la Naturaleza [89]. Una mimesis, por supuesto, en nada ajena a la contraposición Ser-Ente en Heidegger y directamente relacionada con su crítica de la tecnología como culminación para él de la Metafísica [90]. Pero sobre todo una mimesis conflictiva, cuyo transcurso y final necesariamente conllevan el ejercicio de la violencia y de la destrucción por parte del que imita. Esa es la obra criminal debida a la *"instrumentellen vernunft"* a la que Horkheimer [91] con gran acierto se refiere, en concordancia con Marcuse [92] y con Jonas [93], ambos los dos discípulos de Heidegger.

9. Recusación de la Tecnología

(9.1) Pero concluyamos. Bien consta que la lengua de Adán no se ha podido reconstruir. Por lo que se refiere a su vida, se sigue sabiendo lo que se sabía. Así, que, tras ser expulsado con Eva del Paraíso, tuvo un hijo pastor de nombre Abel al que mató Cain su hermano labrador, constatando que Jahvé apreciaba más las ofrendas que le hacía Abel que las que le hacía él y a pesar de ser mucho más trabajosas de lograr. Yahvé reprendió a Caín por su crimen y le condenó a vagar errante por el mundo y a seguir cultivando la tierra, pero sin que por ello le fuera a ser fecunda. Atendiendo a sus peticiones, lo marcó para que nadie atentara contra él en su condición de errante, suscribiendo, de este modo, una alianza. Al fin y al cabo, la agricultura, no por instrumental, deja de ser litúrgicamente cultual y oferente como su nombre evoca. Caín fundó una ciudad y tuvo por hijo a Henoc. Sus hijos, salvo Yabal antepasado de los pastores nómadas, fueron forjadores de metales o músicos. Configuraron, por tanto, al mundo como Boyle quería; esto es, no sólo para gloria del gran Autor de la Naturaleza sino también procurando que en ese mundo fuera posible *"the confort of mankind"*. Del otro hijo que tuvo Caín y de nombre Set descendería Noé. Yahvé, viendo que la violencia se había apoderado de los hombres, decidió aniquilarlos con el Diluvio. Arrepintiéndose, Yahvé salvó a Noé de la progenie de Set que no de la de Henoc y suscribió con él una nueva Alianza, no otra sino la que que el arco iris saludaría. Sostiene Heidegger que el hombre, en su condición de arrojado (*geworfen*) —no del Paraíso ni errando por el mundo—, sino a la existencia, «*es el pastor del Ser*» y que «*el lenguaje es la casa del Ser, habitando en su morada el hombre*» [94] que sólo encontrará la verdad en el *"lichtung"*, en el claro de un bosque. Siendo así, cabe preguntar si, por haberse de nuevo el mundo llenado de constructores de casas, forjadores de metales, músicos instrumentales y cosmopolitas errantes complacidos en sí mismos sin sentido de trascendencia alguno e inmersos en violencia permanente, no habría que repensar de nuevo ese "doble" tecnológico que, en su descarrío y amenazando con un nuevo Diluvio, ha construido en *"mimesis conflictiva"* la razón instrumental, sacrílegamente profanando bosques y espesuras sin respetar noche de verano alguna que en canto de criaturas de la Tierra al Cielo azul se dirigiese. Quizás en algo parecido pudiera haber pensado Heidegger cuando confió a Der Spiegel una entrevista [95] para ser publicada tras su muerte y en la que, obviamente en términos no necesariamente teístas, vino en afirmar: «*Nur noch ein Gott kann uns retten*».

Post scripta a modo de pascalianas "pensées"

Musas, sirenas y artefactos

(i) Clío ya no es la musa de la Historia. Por su parte y con no pocas dificultades, los matemáticos lograron seguir teniendo a Urania como musa, hasta que las Matemáticas se distanciaron de la Astronomía tras caer ésta en manos de la Física moderna. Arquímedes célibe, que era también astrónomo, sufría, según Plutarco, el encantamiento por parte de una sirena. La coyunda monstruosa de Arquímedes con esta sirena pobló el mundo de verdades matemáticas que aún perduran. En sus tiempos libres, Arquímedes se dedicaba también a entretenerse ideando máquinas y artilugios, pero sin el auxilio de una Física propiamente matematizada. Mucho más tarde, vino Euler matemático casto y desmesuradamente casado —los extremos se tocan— con trece hijos debidos a su paternidad. Euler no entendía ni sabía de otra cosa, como lo confesó poco antes de morir, que no fuesen las Matemáticas y sus aplicaciones. Se ignora qué sirena le hechizaba, porque haberla tenía que haberla. Su coyunda con ella por poco deja pequeño al número de resultados matemáticos debidos a Arquímedes.

Los artefactos conceptuales que surgen del acostarse con las sirenas se los denominó, estando todavía Urania de musa de los matemáticos, teoremas; esto es, entes que tienen que ver con la contemplación. Sucedía con ellos, pues, como con la Historia, ya que también su nombre deriva de contemplar. Todo esto, por carecer de interés tecnológico y evocar al pasado, ya está olvidado.

(ii) Si los matemáticos quieren que sus teoremas no sean artefactos maquinalmente engendrados, han de volver a implorar a Urania. Evitarán, en consecuencia, la *"barbarie della riflessione"* denunciada por Vico en su *"Szienza nuova"*. Deberán, pues, no olvidar a Pascal, muy entendido en lo que a ellos matemáticamente les entretiene. Pues Pascal —ser desgarrado y gran teólogo—, vino en enunciar en sus *"Pensées"* la dicotomía *"esprit de finesse, esprit de la géométrie"*. Sucede, además, que el humanismo, por fortuna asistido por las musas, siempre tributó de la *"finesse"*, de la *"fineza del sentir"* que, para Fray Luis de León es propia, no de la ciudad sino *"del campo y de la soledad"*.

(iii) Manifiestamente un artefacto no invita al tacto y menos aún soporta ocultarse tras un velo. Es, cual marioneta de Kleist en danza de Schlemmer, un monstruo que sólo puede bailar desnudo. De modo discreto, además, que no continuo, como corresponde, no a su desnudez, sino a su naturaleza maquinal. Es este monstruo, por exhibicionista y en su impotencia radical, incapaz de ser tocado, capaz como es hasta de electrocutar. Coluccio Salutati en *"De laboribus Herculis"* y tal Ernesto Grassi una y otra vez recuerda, sostiene una concepción de *"scientia"* por definición "'inspirada en las musas" —las musas no eran feministas— para la que la veladura, el velo en su relación con el *"lichtung"* de Heidegger resulta esencial.

Sobre las relaciones de la Geometría Hiperbólica con el Arte

(iv) Constituye todo un lugar común afirmar que el arte islámico —excluidas ciertas manifestaciones en ámbito chiita persa de origen sasánida—, es geometrizante y abstracto. Algo que no debe de extrañar —sobre todo en ámbito de arte sacro—, ya que, para la religión mahometana, Alá es, sin más, lo absolutamente Otro, razón por la que en el Islam y a diferencia del Catolicismo, no ha lugar a *"analogía entis"* alguna entre la criatura y la Divinidad. Un ejemplo paradigmático de este carácter abstracto viene dado por las geométricas azulejerías de la Alhambra de Granada en la que se inspira, como es sabido, buena parte de la obra gráfica de Escher, artista que llegó a colaborar con matemáticos de la talla de George Pólya o Coxeter. Consta asimismo la existencia de estudios psicológicos que establecen relaciones entre el mundo de la esquizofrenia, los grabados de Escher y el universo mental de Enmanuel Swedenborg (1688–1772); es decir, con el de ese conocido científico luterano, significativamente muy interesado en la metalurgia y que acabó para algunos de místico y para otros de esquizofrénico.

(v) Puede afirmarse que algunos artistas occidentales modernos no recusan el tema de la mística en sus "instalaciones", a veces incluso evocando tradiciones místicas islámicas. Es asimismo sabido que países islámicos, como el Emirato de Qatar y otros, acogen, para su exposición, "instalaciones"sobremanera abstractas. Estos hechos no son, en principio, ajenos al Escher fascinado por las azulejerías de la Alhambra granadina y al presunto místico Enmanuel Swedenborg con él psicológicamente emparentado.

(vi) Reforzar la presunta ecuación Escher = Enmanuel Swedenborg resulta, entonces, de lo más sugerente. Es lo que se hace, valiéndose de técnicas matemáticas, en el artículo [96], que contiene, entre otras, las siguientes afirmaciones:

> «*We have demonstrated a representation of all possible 3-dimensional spaces in terms of Escher's Heaven and Hell. There is an amusing parallel of such a representation with the world view of the scientist-turned-mystic, Emanuel Swedenborg (1688–1772).*»

> «*Although Escher, in his writings, makes no mention of Swedenborg, it is conceivable that this Fuchsian group of symmetries was chosen in order to depict three levels each of Heaven and Hell, represented by the six figures in each hexagon. It is likely that Escher was aware of the work of William Blake, who in turn was a "solitary Swedenborgian".*»

(vii) En algunas de las "instalaciones" citadas supra, se cuida meticulosamente la pureza de los metales, la sílice y otras materias empleadas en ellas, materias no precisamente dadas a ser tocadas, como es el caso del mercurio. Se está, de nuevo, en concordancia con el universo de Enmanuel Swedenborg, cuyo interés por los metales ya se ha señalado, sin olvidar a Novalis, otro conocedor de la noche de las minas, en concreto de las de sal. En la medida que la presunta ecuación Escher = Swedenborg sea cierta, dichas "instalaciones" permitirían, pues, situar en un contexto más general el caso Escher [97].

Hágase en mí según tu palabra

(viii) No son ya catasterismos estrellados los que irisan la piel tersa y profunda de ese Cosmos resuelto en esferas al que el Timeo invoca, sino tatuajes que el entrecruzar de las órbitas de los satélites artificiales en los cielos rasga. Configúrase, así, toda una malla de redes así llamadas atesorando, para algunos, el nuevo "saber" de la humanidad denominada. Que, en materia de conocer, el ser humano haya mirado al Cielo, dista de ser nuevo, si bien tenga el modo actual poco o nada que ver con el antiguo. Un mirar este último, dirigido como hacia una esfera que desbordaba y al tiempo acogía, evocando en todo a la propia de un niño recién nacido abriéndose al mundo desde el regazo materno. Un mirar, pues, similar al que otrora se dirigía a una biblioteca, ella colmada de estantes con constelaciones de libros en parpadeo lejano solicitando de nuestra mirada el ser acogidos por nuestras manos, sus páginas sintiendo el tocar de las yemas de nuestros dedos. Un mirar, pues, ajeno a las líneas del punto y raya binario. Líneas éstas que por completo serían ajenas a las esferas, si no fuese porque, similares a ellas y sexagesimalmente escritas, fueron las que proporcionaron información bastante para que —vía saros y exiligmos en su condición de mínimos comunes múltiplos—, pudieran los griegos pitagóricos, si hemos de creer a Censorino de cuando Caracalla, gestar la concepción pitagórica del Cosmos en términos de esferas, haciendo posible no sólo el Timeo de Platón sino el libro XII de la Metafísica de Aristóteles.

(ix) Desarrollos conceptuales que vienen de Alejandro de Afrodisia partiendo de las enseñanzas de su maestro Aristokles y en ámbito aristotélico, vinieron en concluir, después de distinguir en el intelecto de cada ser humano tres tipos diferentes de intelectos —el material, el que "intelige" y el agente—, que este último no pertenece a cada una de las almas individuales humanas, sino que es compartido por todas ellas. De hecho, Alejandro de Afrodisia lo identificó con la "substancia eterna e inmóvil y separada de las cosas sensibles" a la que se refiere Aristóteles en el libro XII, 7 de su Metafísica. Tras tenerla Plotino por una entidad trascendente emanada del Intelecto cósmico, éste a su vez emanado del Uno, correspondió a Alfarabi el ser el primero en identificar al intelecto agente con una esfera alrededor de la Tierra, situada por debajo de la de la Luna y de los demás planetas. Estos desarrollos conceptuales fueron asumidos en gran medida por Avicena y luego por Averroes, pero de forma pretendiendo ser aristotélicamente radical. Así, Averroes rechazó y a diferencia de Avicena, la inmortalidad de las almas individuales que, bajo determinados supuestos, Alfarabi sí que contemplaba. Por eso, sólo el Intelecto agente, común a todos los seres humanos, era, para Averroes y todos los que le siguieron —no otros sino los averroistas latinos incluido Pomponazzi en el siglo XVI—, el único inmortal. Siendo esta la situación, se comprende que no tenga nada de novedoso que Michio Kaku (Harvard University) escriba en su libro "El futuro de nuestra mente" que:

> «*Esto significa que en el futuro algunas personas podrán vivir después de muertas. Habrá bibliotecas de almas. Cuando vas hoy a una biblioteca, ves imágenes y películas de gente que está muerta, pero viven en la librería. En el futuro, tu personalidad y tus memorias estarán almacenadas en un disco y tu tataratataranieto podrá hablar contigo. Ahora puedes leer un libro de una persona muerta, pero entonces podrás hablar con el programa de ordenador que contiene sus memorias. Y podremos ir más allá. ¿Por qué no mezclar el genoma y el conectoma y construir un robot? Serás algo más que un programa de ordenador, volverás a la vida. Es la inmortalidad.*»

Texto falto de novedad, pero estremecedor. Y es que, a su luz, ese "Cosmos vivo y divino", resuelto en esferas bellas, el magno Cielo estrellado su piel tersa y profunda, toda ella en catasterismos irisada, proclamando las cuitas y amores entre humanos, ninfas y deidades" tantas veces evocado, conocería, no las cuitas y amores entre humanos, ninfas y deidades, sino las monstruosas coyundas entre humanos y "robots" denominados generando ni se sabe qué seres. En otros términos: Poussin habría sido asesinado. Entretanto, ya sí que tiene sentido desgraciadamente enunciar, visto lo que se puede hacer con el genoma y el conectoma: «*Hágase en mí según tu palabra, oh Intelecto agente que estás en los Cielos de Internet*».

Lejanas estrellas de la Osa

(x) De Pentesilea reina de las amazonas sólo se acuerda el Cielo dándole su nombre a un asteroide. Si, en cambio, de Aquiles, algunas de cuyas víctimas en Troya —es el caso de Cignus— están incluso catasterizadas. No ha de sorprender que así sea, pues según recoge Diodoro, más sucintamente Proclo y sin que Kleist deforme sus testimonios, Aquiles mató a Pentesilea tras que acudiese en ayuda de los troyanos, huyendo de su patria después de haber perpetrado un asesinato en su familia. Si pues Troya es Oriente, Pentesilea es una de sus valedoras, lo mismo giróvagamente desleída en abstractas "instalaciones" o fascinada por las azulejerías al gusto de Escher.

(xi) Que el Cielo no proclame a Pentesilea en catesterismo es asunto que no debiera de sorprender, pues, como quedó advertido, simboliza ella al Oriente. Además, ya ni existen constelaciones sino tatuajes, no otros sino los que el entrecruzar de las redes sesga, dando lugar a nodos que no a estrellas. Unos nodos que, a los aficionados a Erdös [98] y Barabási, les hacen pensar en los nudos y en los enlaces que el operar háptico del homo faber hace posible en el atar de las cuerdas y que a la Topología invocan.

(xii) No es lo mismo fascinar que encantar. Así, los problemas fascinan y los teoremas encantan, como sus respectivas etimologías ponen de manifiesto y el comentario de Aristófanes recuerda. Por eso, los problemas son más propios de la Esfinge con sus enigmas y asesinando a los que no los resolvieran, hasta que Edipo sí que fue capaz y despechada ella se suicidó o Edipo la mató. Es, entonces, de lo más revelador que, donde el arte de Europa acogió primeramente a las esfinges, fue en los *arabescos* decorativos; es decir, antes de que las porcelanas y las estatuas reparasen en ellas en los jardines del siglo XVIII. La muerte emparenta a Pentesilea y la Esfinge.

(xiii) Encantar de nuevo al mundo es lo que T. J. Clark propone como misión del Arte-Moderno (videoteca Museo Reina Sofía, Madrid). No por cierto desde la tecnología, ya que, *en vez de reencantar al mundo, lo que se haría es acentuar su presente carácter de parque tecnológico con atracciones.* Aducir en contra que se está en el marco de una "sociedad de masas" y que al Arte no le queda otra que el circo, manifiestamente no es de recibo, pues basta recordar que, cuando Plotino, se vivía también en cosmopolitismo megapolita de masas, asimismo complacidas en circo. No obstante, surgió el Neoplatonismo en el que, por cierto, las Matemáticas tenían un muy destacado lugar, asociadas a un tipo de espiritualidad panteista siempre renaciendo en *reencantamiento* del mundo.

(xiv) «*Vaghe stelle dell'Orsa, io non credea/ tornare ancor per uso a contemplarvi/ sul paterno giardino scintillanti*», escribió Leopardi. Es ya, sin embargo, la situación tan desesperada que hasta ellas han sido "instaladas", ni que decir tiene que por los suelos (South London Gallery 1994) y como caídas del Cielo a modo de montoncillos de carbonilla negra, tal si fueran signos anunciando el fin del mundo (Apocalipsis 6/12). Retablo de las maravillas por los suelos o rey desnudo afirmando que porta el más esplendoroso de los vestidos, cabe preguntar entonces. ¿Es que no hay niño —«*Wo Kinder sind, da ist ein goldnes Zeitalter*» afimaba Novalis—, que grite ¡Pero si el rey va desnudo! Sólo hay "Humo" y encima ni se ve. Quizás sí que se vio, pues las estrellas son de carbonilla y lo mismo la recogió en Sierra Morena una pobre niña piconera.

Referencias

[1] Los ejemplos de Arquitas, Arquímedes o Papus no permiten sustentar que la tecnología de la Antigüedad tuviera una fundamentación matemática.

[2] Il saggiatore / Galileo Galilei; edizione critica e commento a cura di Ottavio Besomi e Mario Helbing Pubblicaccione, Roma, Padova: Antenore, 2005.

[3] Aarsleff, H.: American Philosophical Quarterly I (1964), 165–188.

[4] Meinhold, P.: Luthers Sprachphilosophie, Berlin, Lutherisches Verlagshaus, 1958.

[5] Robinson, A.: Non-Standard analysis, Amsterdam, North-Holland, 1974.

[6] Coudert, A.: The impact of the Kabbalah in the seventeenth century, Brill, Leiden, 1999.

[7] Newton, I.: Philosophiae Naturalis Principia Mathematica: the Third edition (1726) with variant readings, assembled and ed. by Alexandre Koyré and I. Bernard Cohen with the assistance of Anne Whitman, Cambridge MA, Harvard UP, 1972.

[8] Gay, P.: Essays in the French Enlightenment, Knopf, New York, 1964, citado por Hine (Referencia 52).

[9] Aarsleff, H.: American Philosophical Quarterly I (1964), 165–188.

[10] Koyré, A.: Newton, Galilée et Platon, Actas del IX Congreso Internacional de Historia de las Ciencias, Madrid, 1959.

[11] Koyré, A.: Etudes newtoniennes, Paris, Gallimard, 1968.

[12] Goethe, W.: Werke : Hamburger Ausgabe. Band 13, Naturwissenschaftliche Schriften I, München, Deutscher Taschenbuch, 1988.

[13] Schiller, F.: Sämtliche Werke. Band 3, Gedichte, München, Artemis und Winkler, 1991.

[14] Novalis: Hymnen an die Nacht, Novalis: Schriften. Die Werke Friedrich von Hardenbergs, P. Kluckholm, R. Samuel, Stuutgart 1977.

[15] Novalis. Citado por Safransky [16].

[16] Safranski, R.: Schiller: Oder Die Erfindung des Deutschen Idealismus, Deutscher Taschenbuch Verlag, München, 2007.

[17] Carone, R.: Cambridge University Press, New York 2005, Plato's Cosmology and its Ethical Dimensions.

[18] West, M. L.: The Cosmology of Hippocrates. De Hebdomadibus, The Classical Quarterly, New Series, Vol. 21, nº 2 (Nov. 1971), págs. 365–388, Cambridge University Press, (`www.jstor.org`)

[19] Broadie, S.: Nature and Divinity in Plato's Timaeus, Cambridge University Press, New York, 2012. Asimismo, Lasserre, F.: Die Fragmente Eudoxus von Knidos, W. de Gruyter, Berlin, 1966.

[20] Tardieu, M.: Sabiens Coraniques et "Sabiens" de Harrán, Journal Asiatique, Tome CCLXXIV, Sociéte Asiatique, Paris 1986.

[21] Dronke, P.: The Spell of Calcidius, Sismel-Edicioni del Galluzo, Firenze 2008.

[22] Adorno en su estancia en USA (1951) se pronunció en los medios de comunicación contra los horóscopos. Ver Müller-Doohm, S.: Adorno, Eine Biographie, Shurkamp Verlag Frankfurt am Main, 2003.

[23] Ver Müller-Doohm, S.: Adorno, Eine Biographie, Shurkamp Verlag Frankfurt am Main, 2003. Adorno acusaba a Heidegger de haber regresado a lo ahistórico y a lo arcaico, cuestionando su modo razonar en Sein und Zeit, Holzwege y Briefe über Humanismus. Por ejemplo, cuando se refiere al hombre teniéndolo por el "pastor del Ser". Como es sabido, Heidegger fue discípulo de Husserl el fundador de la Fenomenología, que le influyó grandemente. Si hemos de creer al Deleuze de 1962, la Fenomenología no es sino la Escolástica del siglo XX. Consecuentemente, el movimiento estudiantil coetáneo con el Wiener Aktionismus de hacia 1965 estaba ideológicamente mucho más enfrentado con Heidegger en su cabaña de Merskirchen que con Adorno conferenciante en Viena. Mas por estar en Viena justo meses antes de morir, fue Adorno el atacado por las estudiantes. En efecto, unas universitarias vienesas, mostrándose en ropa interior y en un episodio que Hans Kelsen teórico del Estado de Derecho tendría por propio de "violencia de género" femenina, se fueron a por Adorno conferenciante acosándolo y propinándole el disgusto que es de imaginar. Sobre Deleuze y la Fenomenología como Escolástica del siglo XX, ver Alain Beaulieu: L'incarnation phénoménologique à l'épreuve du 'corps sans organes. Laval theologique et philosophique, Vol 60, nº 2, 2004, `http://id.erudit.org/iderudit/010348ar`.

[24] Pöltner, G.: El lugar de la música en el pensamiento de Heidegger, Thémata Revista de Filosofía, nº 34, 2005.

[25] Révész, G.: Talent und Genie, Lehnen, München 1952. También: Weininger, O.: Geschlecht und Charakter, Braumüller, 1923 (1ª edición 1903).

[26] Révész, G.: Einführung in die Musikpsychologie, Amsterdam, 1946.

[27] Arnold, V. I.: On teaching Mathematics, Palais de la Découverte (7–March–1997) (traducción de Goryunov, A.V.).

[28] Fortea Pérez, J.: Las Matemáticas no son una cuestión de tacto, Madrid, autoedición (ISBN nº 84-920791-2-6), 1995

[29] Panofsky, E.: Galileo as a critic of the arts, M. Nijhoff, The Hague,1954.

[30] Bergson, H.: Essai sur les données immediates de la conscience, Oeuvres, PUF, Paris, 1959. Cuestión estudiada por Aron Gurwitsch en su obra The Field of Consciousness, Duquesne University Press, Pittsburg, 1964.

[31] van der Waerden, L.: Science Awakening, Noordhoff International, cop.: Leyden, 1975.

[32] Baudelaire, Ch.: Les Fleurs du mal, Gallimard, Paris, 1978.

[33] Fortea Pérez, J.: Introducción a los procesos diferenciales, Tecnos, Madrid 1981.

[34] Rosen, S.: Thought and Touch: A note on Aristotle's De Anima, Phronesis, 6, 127–137, 1961.

[35] Mueller, I.: Aristotle on Geometrical objets, Archiv für Gesischte der philosophie, 52, 156–171, 1970.

[36] de Haas, F.: John Philoponus' new definition of prime matter : aspects of its background in Neoplatonism and the ancient commentary tradition, E.J. Brill, Leiden, New York, Köln, 1997.

[37] Ver referencia 27. También: Mueller, I.: Aristotle's doctrine of abstraction in the commentators, en: Aristotle Transformed edited by Richard Sorabji, Cornell University Press, 1990.

[38] Révész, G.: Psychology and Art of the Blind, Longmans, London, New York, Toronto, 1950.

[39] Poincaré, H.: Analisis Situs, Journal de l'École Polytechnique, Paris, 1895.

[40] Dieudonné, J.: Foundations of Modern Analysis, New York, Academic Press, 1960.

[41] Montesinos Amilibia, J.: Números, Combinatoria y nudos: de lo discreto a lo continuo, Real Academia de Ciencias Exactas, Físicas y Naturales, Madrid, 1996. El autor en la página 4 y refiriéndose a esos geómetras denominados "topólogos de baja dimensión", manifiesta que «sólo sab(en) mover las manos».

[42] Herder, J.: Kritisches Wälder 4, Aufbau-Verlag, Berlín, 1990.

[43] Catálogo de la exposición: The Sacred Made Real. Spanish Painting and Sculpture 1600–1700, National Gallery (21 October 2009 – 24 January 2010), London.

[44] Riegl, A.: Spätrömische Kuntsindustrie, Oesterr. Archaeol. Institut, Wien, 1927.

[45] Berenson, B.: Aesthetics and History, Constable Publishers, london 1950. La edición en italiano es de 1948.

[46] Cortadura que es un pulsar de la cuerda. Pulsar que si se quisiese armonioso, buscaría que todo punto quedase integrado en acorde. Mas no es así, ya que entre dos números en razón epimórica no hay media proporcional y ese es el caso de raíz de dos en el monocordio de longitud dos.

[47] Aristóteles: Metafísica III,2,998a,3, Edición García Yebra, Gredos, Madrid, 1970.

[48] Gurwitsch, A.: The Field of Consciousness, Duquesne University Press, Pittsburg, 1964.

[49] Aristóteles: De Sensu 3,439a30. Kirk, G., Raven, J.: The Presocratic Philosophers. A critical History with a selection of texts, Cambridge University Press, 1969.

[50] Heidsieck, F.: Henri Bergson et la Notion d'Espace. PUF, Paris, 1961.

[51] Giroux, L.: Durée pure et temporalité, Bergson et Heidegger.

[52] Merleau-Ponty, M.: Phénoménologie de la perception, Gallimard, Paris, 1945. También y esencialmente [41].

[53] Gouhier, H.: Les Conversions de Maine de Biran, Libraire Philosophique J. Vrin, Paris 1948. En la pag. 44 informa en relación con el proyecto que tenía Rousseau de escribir "La morale sensitive ou le matérialisme du sage", empresa que no llevo a término.

[54] Condillac, E.: Traité des sensations, Oeuvres philosophiques, edición e introducción a cargo de Georges Le Roy, PUF. Paris, 1947.

[55] Piaget, J., Beth, W.: Epistemologie mathématique et psychologie. Essai sur les relations entre la logique formelle et la pensée reelle, PUF, París 1961.

[56] Bédarida, H.: Les premiers Bourbons de Parme et l'Espagne (1731-1808), Librairie Ancienne Honoré, Champion, Paris, 1928.

[57] Die philosophischen Schriften von Gottfried Wilhelm Leibniz, herausgegeben von C.I. Gerhardt, Band 2 (Leibniz-Arnauld Briefen), Georg Olms, Hildesheim, 1960–1961.

[58] Arnauld, A., Nicole, P.: La Logique ou l´Art de penser, edición e introducción a cargo de Louis Marin, Flammarion, Paris, 1970.

[59] Aarsleff, H.: The History of Linguistics and Professor Chomsky, Language 46, 570–585.

[60] McNiven Hine, E.: A Critical Study of Condillac's Traité des Systemes, Martinus Nijhoff Publishers, The Hague/ Boston7 London, 1979.

[61] Gilson, E.: The unity of philosophical experience, Charles Scribner'sons, New York, 1937

[62] Diderot, D.: Lettre sur les aveugles, edition critique par Robert Niklaus, Droz, Géneve, Paris, Minard, 1970.

[63] Romanelli, P.: The Philosophy of Giovanni Gentile, S. F. Vanni Inc.: New York, 1938.

[64] Szabó, A.: Anfange der griechischen Mathematik, Akadémiai Kiado, Budapest, 1969.

[65] Scheler, M.: Die Stellung des Menschen im Kosmos, Otto Reichl, Darmstadt, 1928.

[66] Kafka, F.: Ein Bericht für eine Akademie / Die Verwandlung (1917), Ernst Klett Verlag, Stuttgart, 2011.

[67] Wells, G.: The origin of Language, Open Court, Chicago and Lasalle, 1987. Más próximo a la tesis que se ha expuesto: Révész, G.: Ursprung und Vorgeschichte der Sprache, Francke Verlag, Bern, 1946.

[68] Gouhier, H.: Maine de Biran par lui-même, Les éditions du seuil, Paris, 1970.

[69] Maine de Biran, M.: Essai sur les Fondements de la psychologie, Oeuvres de Maine de Biran, accompagnées de notes et d'appendices (ed. Tisserand), F. Alcan, París, 1920.

[70] Hadot, P.: Plotin ou la simplicité du regard, Gallimard, 77–78, Paris, 1997.

[71] Worringer, W.: Abstraktion und Einfülung, Piper&Co. Verlag, München, 1908.

[72] Ver [70]. En la pág. 77 del libro de P. Hadot se indica que la referencia a Leonardo se debe a M. Louis Frank.

[73] Thibaudet, J.-Y.: Le bergsonisme, Editions de la Nouvelle Revue Française, Paris,1924.

[74] Gouhier, H.: Etudes sur l'Histoire des idées en France depuis le XVII siècle (chs. V y VI), Librairie philosophique J. Vrin, Paris, 1980.

[75] Vico, G.: La scienza nuova 1730 a cura de Paolo Cristofolini, Guida Editore, Napoli, 2004. También: Berlin, I.: Vico and Herder: Two Studies in the History of Ideas, Chatto windus, London, 1980.

[76] Aristóteles: Metafísica XIII 6, 1080b16. Kirk, G., Raven, J.: The Presocratic Philosophers. A critical History with a selection of texts, Cambridge University Press, 1969.

[77] Capek, M.: La théorie bergsonienne de la matiére et la physique moderne, Revue philosophique de la France et de l' Étranger, t. CXLIII, n° 1-3, 1953.

[78] Ver [69]. También: The concepts of space and time : their structure and their development. Edited by Milic Capek, Dordrecht, D.Reidel, 1976

[79] de Broglie, L.: Phisique et microphisique, Albin Michel, Paris, 1947.

[80] Bohm, D.: Causality and Chance in in Modern Physics, prólogo de Louis de Broglie, Routledge&kegan Paul Ltd.: London, 1957.

[81] Eddington, A.: New York, The Macmillan Company; Cambridge University, Press, 1928. Citado por Capek. Ver también [69].

[82] Whitehead, A.: Process and Reality: An Essay in Cosmology, corrected edition (1979) edited by David Ray Griffin and Donald W. Sherburne, Free Press, 1929.

[83] Bergson, H.: Introduction à la Métaphisique (La Pensée et le Mouvant), PUF, Paris, 1959.

[84] Thom, R.: Stabilité structurelle et morphogénése : essai d'une théorie générale des modéles (pág. 166), Reading, Massachusetts : W. A. Benjamin, 1972.

[85] Ver [73]. También y esencialmente Bergson, H.: L'Évolution créatrice, PUF, París, 1959.

[86] Ver Postcripta.

[87] Barthélemy-Madaule, M.: Bergson et Teilhard de Chardin (pág 185), Éditions du Seuil, Paris, 1963.

[88] von Balthasar, U.: Kosmische Liturgie: Das Weltbild Máximus' des Bekenners, Johannes Verlag, Einsiedeln, 1988.

[89] Girard, R.: La Violence et le Sacré, Grasset, Paris, 1972.

[90] Heidegger, M.: Die Frage nach der Technik, Gesamtausgabe, Klostermann, Frankfurt am Main, 2001.

[91] Horkheimer, M.: Zur Kritik der instrumentellen Vernunft, S. Fischer, Frankfurt am Main,1967.

[92] Marcuse, H.: One Dimensional Man, Routledge&Kegan Paul Ltd, London, 1964.

[93] Jonas, H.: Das Prinzip Verantwortung: versuch einer ethik für die technologische Zivilisation, Suhrkamp, Frankfurt am Main, 1984. También: Wolin, R.: Heidegger's children: Hannah Arendt, Karl Löwith, Hans Jonas, and Herbert Marcuse, Princeton University Press, 2001.

[94] Heidegger, M.: Brief über Humanismus, Vittorio Klostermann, Frankfurt am Main, 1968.

[95] Heidegger, M.: Entrevista en Der Spiegel, 31-mayo-1976. Traducción al español de la frase «Nur noch ein Gott kann uns retten»: Solamente un Dios puede todavía salvarnos.

[96] Aitchison, I. R., Rubinstein, J. H.: Heaven&Hell, Mathematics Department, University of Melbourne. Internet.

[97] El autor de este escrito prepara un trabajo sobre el tema expuesto. Sirva esta nota para adelantar algunas reflexiones complementarias a las ya expuestas. El tema de las minas va asociado a una fascinación por las cuevas, galerías y minas abandonadas. Una fascinación que se podría tipificar aún más valiéndose de las categorías psicoanalíticas de Melanie Klein con tal de apreciar una fase "esquizo-paranoide" *que se complace* en dentelladas o agresiones "orales" contra tierras y montañas por completo inocentes, complacencia a la que seguiría una fase "depresivo-*reparadora*", en esta ocasión con los mineros de agraciados. En efecto, dentelladas similares a las indicadas, a cargo de grúas y excavadoras, se contemplaron con gran gozo mientras se realizaban hacia 1980 en unas grandes grandes obras nada ecológicas en Madrid y sobre las que se declaró: «Y me concienсé del valor como espectáculo de los movimientos de terreno: ese gran volumen de tierra, esa mezcla de lo mecánico y lo natural... Era una maravilla». Es más, a la pregunta de si una extracción de tierras semeja una intrusión del estado natural por intrusión de lo mecánico, se contestó «No lo veo así» ("Una obra para un espacio", libreto publicado en 1987 por la Comunidad de Madrid). En resumen: el estado de cosas descrito lleva de bruces a Deleuze, Guattari, Blanchot o Artaud, los dos primeros apelando, a la *démétaphorisation* y al *"flujo"* de la materia entendido como "metálico". En suma, se acaba en el Antiedipo de Deleuze-Guattari con parada de nuevo en el mundo de la esquizofrenia. Peor: como las dentelladas generaron víctimas, nada más indicado que los libros de títulos "The seven Rites of the Lakota" y "Trauma and Resilence in American Indian and African Southers History", los dos debidos a Ulrike Wiethaus (1996), curiosamente autora también del libro "Ecstatic Transformation or Transpersonal Pschiology in the work of Mechtild of Magdeburg" tratando del tema de la piedra o la montaña identificada con la Deidad.

[98] Hoffman, P.: The man who loved only numbers: the story of Paul Erdös and the search for mathematical truth, Fourth State, London 1998.

Persistencia uniforme de atractores

Antonio Giraldo, Victor Fernández Laguna
y José M. Rodríguez Sanjurjo

Departamento de Matemática Aplicada a las Tecnologías de la Información y las Comunicaciones
Universidad Politécnica de Madrid, 28040 Madrid, Spain
antonio.giraldo@upm.es

Departamento de Matemáticas Fundamentales, Facultad de Ciencias
Universidad Nacional de Educación a Distancia, 28040 Madrid, Spain
vfernan@mat.uned.es

Departamento de Geometría y Topología, Facultad de CC Matemáticas
Universidad Complutense de Madrid, 28040 Madrid, Spain
jose_sanjurjo@mat.ucm.es

Dedicado al profesor José María Montesinos.

Resumen

En este trabajo expositivo consideramos familias parametrizadas de flujos en espacios metrizables localmente compactos y damos caracterizaciones de aquellas familias parametrizadas de flujos para las que la persistencia continúa.

Por otra parte, estudiamos las bifurcaciones de Poincaré-Andronov-Hopf generalizadas, para familias parametrizadas de flujos en los puntos de la frontera de $\mathbb{R}^n_+$ o, más generalmente, de una variedad n-dimensional, y mostramos que esta clase de bifurcaciones produce una familia completa de atractores que evolucionan desde el punto de bifurcación y tienen interesantes propiedades topológicas. En particular, en algunos casos, la bifurcación transforma un sistema con propiedades extremas de no permanencia en un sistema uniformemente persistente. Estudiamos cuando ocurre este fenómeno y proporcionamos un ejemplo construido combinando una interacción de tipo Holling con una bifurcación horca.

Finalmente, estudiamos la estructura interna del atractor global de un flujo uniformemente persistente. Mostramos que la restricción del flujo al atractor global tiene propiedades de dualidad que pueden ser expresadas en términos de ciertas descomposiciones atractor-repulsor. Estudiamos también algunas descomposiciones de Morse naturales del flujo y calculamos sus ecuaciones de Morse. Estas ecuaciones proporcionan condiciones necesarias y suficientes para la existencia de atractores con la forma de $\mathbb{S}^1$, o tal que sus suspensiones tienen forma esférica.

2010 Mathematics Subject Classification: 37B25, 37C70, 37G35.
Key words: Persistencia uniforme, atractor, repulsor, robustez, continuación, ecuaciones de Morse, forma.

1. Introducción

En este trabajo expositivo se revisan algunas cuestiones relacionadas con la persistencia de flujos de sistemas dinámicos. Es este un tópico clásicamente ligado a la dinámica de poblaciones, cuyo problema central es el de determinar si alguna componente de la población está condenada a la extinción o si, por el contrario, dicha población evoluciona hacia un estado de equilibrio estable en el que se alcanza la coexistencia entre todas las componentes de la población.

El término persistencia se suele aplicar a los sistemas cuyas órbitas no se acercan a la frontera del ortante no negativo $\mathbb{R}^n_+$ cuando $t \to \infty$, situación que implicaría el riesgo de extinción. El problema que se plantea es, por tanto el de determinar condiciones que prevengan la existencia de soluciones que se acerquen a la frontera. Como G. Butler, H.I. Freedman y P. Waltman destacan en [3], este problema es de gran importancia para el modelado de poblaciones biológicas donde tales condiciones eliminan la posibilidad de que una de ellas se acerque arbitrariamente a cero en un modelo determinista, arriesgándose entonces a la extinción en una interpretación realista del modelo.

Uno de los resultados clásicos es el Teorema de Butler-Waltman, que establece condiciones suficientes para la detección de una forma más fuerte de persistencia llamada persistencia uniforme y también cooperación, coexistencia permanente o, sencillamente, permanencia. Esto significa que las semitrayectorias positivas son eventualmente uniformemente acotadas lejos de la frontera (ver [3], [23] y [53]).

La robustez de la persistencia uniforme ha sido investigada por varios autores. Aunque la persistencia uniforme no es una propiedad robusta, se han encontrado condiciones suficientes para su robustez. En este trabajo mostramos algunos resultados en este sentido que utilizan el punto de vista de la teoría del índice de Conley ([6, 7]), en particular la noción de continuación.

Consideraremos familias parametrizadas de flujos disipativos y veremos que la persistencia uniforme tiene propiedades de continuación débil. Una consecuencia de este hecho es que perturbaciones pequeñas del flujo no llevan a la extinción de poblaciones en un determinado rango (que puede ser arbitrariamente escogido). Veremos también una condición de regularidad introducida en [48] que garantiza la continuación.

En la siguiente sección veremos que esta condición suficiente de regularidad ha de ser requerida para todos los puntos $x \in \mathring{E}$, y que no basta si la exigimos sólo para puntos cercanos a los conjuntos de Morse de una descomposición del compacto invariante maximal de ∂E, como en el Teorema de Butler-Waltman & Garay. Una de las motivaciones de los resultados de esta sección es proporcionar una versión de este resultado en el espíritu del Teorema de Butler-Waltman & Garay. En el Teorema 7 y en el Corolario 3 damos condiciones necesarias y suficientes para esta formulación. También identificamos condiciones más débiles que garantizan la continuación de la persistencia uniforme para trayectorias acotadas.

Otra motivación detrás de los resultados de este trabajo es el estudio de la persistencia uniforme en el contexto de las bifurcaciones, en particular aquellas bifurcaciones que surgen por la pérdida de estabilidad en puntos del flujo, tal como las bifurcaciones generalizadas de Poincaré-Andronov-Hopf. En la sección 5 veremos que este tipo de bifurcaciones produce una familia completa de atractores que evolucionan a partir del punto de bifurcación y que tienen interesantes propiedades topológicas. Una posible consecuencia de la bifurcación es un cambio cualitativo en las propiedades de persistencia del sistema. En algunos casos la bifurcación transforma un sistema con propiedades extremas de no permanencia en un sistema uniformemente persistente. Estudiamos cuando sucede este fenómeno y damos un ejemplo construido combinando una interacción de tipo Holling con una bifurcación horca.

La última sección del artículo está dedicada al estudio de la dinámica dentro del atractor global de un flujo uniformemente persistente. La existencia de un punto repulsivo p implica la existencia de un atractor dual A con forma esférica cuya región de atracción es $\operatorname{int}\mathbb{R}^n_+ - \{p\}$. Por *forma* entendemos la noción introducida y estudiada por K. Borsuk en [2] que se ha convertido en una herramienta fundamental para el estudio de propiedades globales de conjuntos compactos invariantes, y de atractores de sistemas dinámicos. Este resultado tiene interesantes connotaciones topológicas ya que el atractor A no es en general homeomorfo a una esfera. Estudiamos también algunas descomposiciones de Morse naturales de flujos uniformemente persistentes, calculando sus ecuaciones de Morse y probando que esas ecuaciones son suficientes para detectar la existencia de atractores con la forma de $\mathbb{S}^1$ en el plano o atractores cuya suspensión tiene la forma de una esfera para casos en dimensiones más altas.

En lo que sigue fijamos alguna terminología y enunciamos algunos resultados que serán usados más adelante. Un atractor de un flujo $\varphi : E \times \mathbb{R} \to E$, donde E es un espacio metrizable localmente compacto, es en este trabajo, un compacto invariante asintóticamente estable. Un repulsor es un compacto invariante negativamente asintóticamente estable, i.e. un atractor para el flujo inverso. Usaremos la siguiente caracterización de los repulsores (ver [41]): Un compacto invariante K es un repulsor si y sólo si existe un entorno U de K en E tal que para todo $x \in U - K$ existe $t > 0$ tal que $xt \notin U$. Existe una caracterización dual para atractores.

El flujo φ se dice disipativo si $\omega(x) \neq \emptyset$ para todo $x \in E$ y $\cup_{x\in E}\omega(x)$ tiene adherencia compacta. Si E no es compacto consideraremos habitualmente la compactificación de Alexandrov $E \cup \{\infty\}$ y el flujo extendido $(E \cup \{\infty\}) \times \mathbb{R} \to E \cup \{\infty\}$ que deja fijo ∞. En estas condiciones ser disipativo equivale a que $\{\infty\}$ sea un repulsor (ver [11] y [18]). Nótese que el atractor dual de $\{\infty\}$ es un atractor global para el flujo φ.

Existe una forma más fuerte de disipatividad para familias de flujos. Si $\varphi_\lambda : E \times \mathbb{R} \longrightarrow E$, $\lambda \in I = [0, 1]$, es una familia parametrizada (continua) de flujos, entonces se dice que φ_λ es uniformemente disipativo si $\omega_\lambda(x) \neq \emptyset$ para todo $x \in E$ y todo $\lambda \in I$ y el conjunto $\Omega = \bigcup_{x\in E,\lambda\in I} \omega_\lambda(x)$ tiene adherencia compacta.

En este trabajo E será generalmente un subconjunto cerrado de X, donde X es un espacio métrico localmente compacto, y denotaremos por ∂E la frontera de E en X. Diremos que el flujo disipativo $\varphi : E \times \mathbb{R} \to E$ es uniformemente persistente si existe un número real $\beta > 0$ tal que $\liminf\{d(\varphi(x, t), \partial E) \mid t \to \infty\} \geq \beta$, para todo

$x \in \mathring{E}$. En todo el trabajo supondremos que ∂E es invariante para el flujo φ. Si E es compacto entonces φ es uniformemente persistente si y sólo si ∂E es un repulsor de φ. Si φ es disipativo y E no es compacto entonces φ es uniformemente persistente si y sólo si $\partial E \cup \{\infty\}$ es un repulsor para el flujo extendido a $\hat{E} = E \cup \{\infty\}$ (ver [11] para demostraciones de estos resultados). En este caso existe un atractor dual K cuya región de atracción es el interior $\mathring{E}$. Llamaremos a K el *atractor global interno*. K no debe ser confundido con el atractor global de φ, que es un conjunto más grande.

Sea K un compacto invariante de φ y $K_0 \subset K$ un subcompacto invariante. Decimos que K_0 es de *tipo repulsor* si $W^s(K_0) \subset K$ donde $W^s(K_0)$ es la variedad estable de K_0. Este tipo de conjuntos fueron introducidos y estudiados por Wójcik (ver [57]).

Proposición 1. *Sea K un compacto invariante de $\varphi : E \times \mathbb{R} \to E$ y $K_0 \subset K$ un subcompacto invariante. Supongamos que K_0 es aislado como subconjunto de E (i.e. existe un entorno aislante de K_0 en E). Entonces $K_0 \subset K$ es de tipo repulsor si y sólo si existe un entorno U de K_0 en E tal que para todo punto $x \in U - K$ existe $t > 0$ tal que $xt \notin U$.*

Utilizamos en este trabajo algunas nociones clásicas de la topología algebraica, en particular las teorías de homología, cohomología y dualidad junto con algunos rudimentos de teoría de la forma. La noción de tipo de homotopía es demasiado rígida para el estudio de los objetos topológicos que aparecen en sistemas dinámicos. Por esta razón, muchos autores han usado en su lugar la teoría de la forma de Borsuk como una herramienta esencial que proporciona una visión geométrica de la estructura global de compactos, principalmente de aquellos con estructura topológica complicada como son muchos atractores y conjuntos invariantes. Para el beneficio del lector presentamos aquí una muy breve introducción, esencialmente basada en la presentación de este tema dada por Kapitanski y Rodnianski en [26].

Un espacio metrizable M es un retracto entorno absoluto ($M \in$ ANR) si para todo homeomorfismo h de M sobre un subconjunto cerrado $h(M)$ de un espacio metrizable X existe un entorno U de $h(M)$ en X tal que $h(M)$ es un retracto de U. Dos importantes caracterizaciones de los ANRs son las siguientes.

Teorema 1. *Un espacio metrizable M es un ANR si y sólo si para toda aplicación continua $f : Y \longrightarrow M$, definida en un subconjunto cerrado Y de un espacio metrizable Y', existe un entorno U de Y en Y' y una aplicación $f' : U \longrightarrow M$, que es una extensión continua de f.*

Teorema 2. *Un espacio metrizable M es un ANR si y sólo si es homeomorfo a un retracto de un subconjunto abierto de un conjunto convexo de un espacio de Banach.*

En particular, los subconjuntos abiertos de los espacios euclídeos son ANRs.

Todos los espacios métricos se pueden considerar como subconjuntos de ANRs. De hecho, por el Teorema de Kuratowski-Wojdyslawski [24], todo espacio métrico puede ser sumergido en un ANR como un subespacio cerrado.

Sean X un subconjunto cerrado de un ANR M e Y un subconjunto cerrado de un ANR N. Denotemos por $\mathbb{U}(X;M)$ (resp. $\mathbb{U}(Y;N)$) el conjunto de todos los entornos abiertos de X en M (resp. Y en N).

Sea $\mathbf{f} = \{f : U \longrightarrow V\}$ una colección de aplicaciones continuas de ciertos entornos $U \in \mathbb{U}(X;M)$ en $V \in \mathbb{U}(Y;N)$. Decimos que $\mathbf{f}$ es una mutación si:

1) Para todo $V \in \mathbb{U}(Y;N)$ existe (al menos) una aplicación $f : U \longrightarrow V$ en $\mathbf{f}$.
2) Si $f : U \longrightarrow V$ está en $\mathbf{f}$ entonces la restricción $f|_{U_1} : U_1 \longrightarrow V_1$ está también en $\mathbf{f}$ para todo entorno $U_1 \subset U$ y todo entorno $V_1 \supset V$.
3) Si dos aplicaciones $f, f' : U \longrightarrow V$ están en $\mathbf{f}$ entonces existe un entorno $U_1 \subset U$ tal que las restricciones $f|_{U_1}$ y $f'|_{U_1}$ son homótopas.

Un ejemplo de mutación es la mutación identidad $\mathbf{id}_{\mathbb{U}(\mathbf{X};\mathbf{M})}$ consistente en las aplicaciones identidad $i : U \longrightarrow U$, para todo entorno $U \in \mathbb{U}(X;M)$. Las nociones de composición de mutaciones y homotopía de mutaciones pueden ser definidas de una manera directa que el lector podrá deducir (los detalles se pueden ver en [26]).

Dos espacios métricos X y Y tienen la misma forma si pueden ser sumergidos como conjuntos cerrados en ANRs M y N, respectivamente, de tal forma que existen mutaciones $\mathbf{f} = \{f : U \longrightarrow V\}$ y $\mathbf{g} = \{g : V \longrightarrow U\}$, tales que las composiciones $\mathbf{gf}$ y $\mathbf{fg}$ son homótopas a las mutaciones identidad $\mathbf{id}_{\mathbb{U}(\mathbf{X};\mathbf{M})}$ e $\mathbf{id}_{\mathbb{U}(\mathbf{Y};\mathbf{M})}$ respectivamente.

La noción de forma de los espacios no depende de los ANRs en que se sumergen ni de las inmersiones escogidas.

Dos espacios con el mismo tipo de homotopía tienen la misma forma.

Dos ANRs tienen la misma forma si y sólo si tienen el mismo tipo de homotopía. Una consecuencia de las dos afirmaciones anteriores es que la noción de forma puede ser vista como una generalización de la noción de tipo de homotopía.

Para un tratamiento completo de la teoría de la forma remitimos al lector a [2, 8, 10, 32, 33]. El uso de la teoría de la forma en sistemas dinámicos se puede ver en los artículos [14, 15, 16, 17, 19, 26, 36, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48]. Para información sobre los aspectos básicos de la teoría de sistemas dinámicos recomendamos [1, 40, 56]. En [22, 23, 53, 57] se puede encontrar información sobre algunos aspectos de la permanencia. En particular, en [57] Wójcik establece interesantes relaciones entre la permanencia y la teoría del índice de Conley. Para descomposiciones de Morse recomendamos [7, 26, 28, 29, 46]. Finalmente, las principales referencias usadas para topología algebraica han sido los libros de Hatcher [20] y Spanier [54].

2. El Teorema de Butler-Waltman

El Teorema de Butler-Waltman [4] es uno de los resultados más relevantes en la teoría de flujos persistentes. Proporciona un criterio para la persistencia uniforme que en las aplicaciones más elementales puede ser reducido a hipótesis comprobables. Este resultado muestra que algunas cuestiones de persistencia pueden ser resueltas imponiendo condiciones adecuadas en el flujo en la frontera. Butler y Waltman enunciaron sus resultados en términos de recubrimientos acíclicos aislados pero más tarde Garay presentó en [11] una reformulación en términos de descomposiciones de Morse. Los resultados de Garay están formulados en el espíritu de la Teoría del índice de Conley y, en particular, hace uso de nociones relacionadas con recurrencias en cadena. En el trabajo [13] hemos creído interesante comenzar la exposición dando una nueva demostración corta del Teorema de Butler-Waltman-Garay, que tiene la ventaja de usar sólo las nociones más elementales de la dinámica topológica y puede ser útil a lectores no familiarizados con la teoría de recurrencias en cadena. Establecemos primero una versión abstracta del Teorema (según fue refraseado por Hofbauer [22]) y después el Teorema propiamente dicho.

Teorema 3. *Sea $\varphi : Y \times \mathbb{R} \to Y$ un flujo en un espacio métrico compacto Y. Supongamos que K es un conjunto compacto invariante y que $\{M_1, \ldots, M_n\}$ es una descomposición de Morse de K, donde todos los M_i son de tipo repulsor y aislados en Y. Entonces K es un repulsor.*

Usando el Teorema anterior, en [13] damos una nueva demostración del Teorema de Butler-Waltman, en la formulación dada por Garay. Denotamos por $W^+(M)$ la variedad estable de M, i.e. el conjunto $W^+(M) = \{x \in E \mid \omega(x) \subset M\}$.

Teorema 4. *Sea X un espacio métrico localmente compacto y sea E un subconjunto cerrado de X. Sea φ un sistema dinámico disipativo en E para el que ∂E es invariante. Sea $\mathcal{M} = \{M_1, M_2, \ldots, M_n\}$ una descomposición de Morse para $\varphi|_M$, donde M es el conjunto compacto invariante maximal en ∂E. Supongamos además que, para cada $i \in \{1, 2, \ldots, n\}$,*

a) existe $\gamma > 0$ tal que el conjunto $\{x \in \mathring{E} \mid d(x, M_i) < \gamma\}$ no contiene trayectorias enteras, y

b) $\mathring{E} \cap W^+(M_i) = \emptyset$.

Entonces φ es uniformemente persistente.

3. Continuación de la persistencia uniforme

Una propiedad deseable para la persistencia uniforme es la robustez, i.e. la preservación de la persistencia uniforme después de que el flujo haya sido sometido a pequeñas perturbaciones. Algunas contribuciones importantes en este campo son las siguientes.

V. Hutson consideró en [25] el sistema $\dot{x} = f(x)$ en $\mathbb{R}^n_+$, donde $f = (f_1, \ldots, f_n)$, $x = (x_1, \ldots, x_n)$, y $\dot{x}_i = dx_i/dt$, y f cumple lo siguiente:

1) f es localmente Lipschitz,
2) para cualquier $x(0) = x$, la solución $x(t)$ por x existe para todo $t \geq 0$,
3) existe $b_0 > 0$ tal que para cualquier $x \in \mathbb{R}^n_+$, $x(t) \in B(b_0)$ para algún $t > 0$, donde $B(b_0)$ es la bola abierta con centro el origen y radio b_0,
4) $f_i(x) = 0$ si $x_i = 0$, para cada $i = 1, \ldots, n$,

junto con un sistema perturbado $\dot{x} = f(x) + g(x, t)$, donde g se comporta lo suficientemente bien como para asegurar la existencia y unicidad de soluciones. Suponiendo que las soluciones del sistema perturbado son *uniformemente acotadas*, es decir, que para cualquier $\alpha > 0$ existe $\beta(\alpha)$ tal que si $t \geq t_0 \geq 0$ y $x_0 \in \bar{B}(\alpha)$ (la bola cerrada), entonces $x(t, t_0, x_0) \in B(\beta(\alpha))$, donde $x(t, t_0, x_0)$ es la solución con $x(t_0, t_0, x_0) = 0$.

Hutson introdujo y estudió una noción de *repulsión bajo perturbaciones* para la frontera $\partial\mathbb{R}^n_+$. Su resultado principal es que si el sistema original $\dot{x} = f(x)$ es uniformemente persistente, entonces la frontera $\partial\mathbb{R}^n_+$ es repulsiva bajo perturbaciones. El lector puede consultar el artículo de Hutson [25] para más detalles.

Schreiber estudió en [52] campos de vectores C^r, $x_i^* = x_i f_i(x)(i = 1, \ldots, n)$, que generan flujos disipativo en $\mathbb{R}^n_+$, obteniendo una condición necesaria y suficiente, para la robustez de clase C^r de la permanencia, en términos de las tasas de crecimiento per-cápita $\int f_i d\mu$ con respecto a medidas invariante μ.

Los resultados de Schreiber fueron mejorados por Hirsch, Smith y Zhao en [21] donde probaron, entre otras cosas, que la persistencia uniforme es estable respecto a una perturbación por un campo de vectores Lipschitz $\mathcal{C}^0-$pequeño. También encontraron condiciones suficientes para que una familia parametrizada φ_λ de semiflujos discretos sea robusta. Básicamente demostraron que la persistencia uniforme de φ_0 y la persistencia débil uniforme de todo φ_λ, uniforme en los parámetros, junto con algunos requisitos asintóticos son suficientes para garantizar la robustez.

Garay y Hofbauer demostraron resultados más generales en [12], donde dieron condiciones suficientes para la robustez de la permanencia siguiendo un enfoque más clásico usando dinámica topológica, en particular funciones de Lyapunov medias "buenas" (GALF), el Teorema de Zubov–Ura–Kimura, y descomposiciones de Morse.

Es fácil ver que, en un contexto general, la persistencia uniforme no es una propiedad robusta. La ilustración de la Figura 1 muestra que pequeñas perturbaciones de un flujo uniformemente persistente pueden destruir esta propiedad.

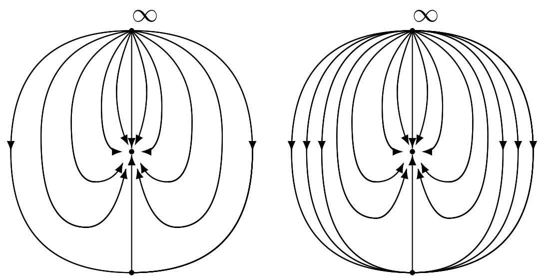


Figura 1: Un flujo uniformemente persistente φ y una pequeña perturbación de φ

La figura de la izquierda muestra el flujo para $\lambda = 0$. La figura de la derecha muestra el flujo para $\lambda > 0$. Para $\lambda = 0$ todas las órbitas son atraídas por el punto en el centro, con la excepción del punto fijo inferior y de las dos órbitas exteriores atraídas por él. Para $\lambda > 0$ hay dos conjuntos de órbitas "paralelas" atraídas por el punto fijo inferior. Estos conjuntos de órbitas encogen según λ se acerca a 0.

En esta sección y la siguiente veremos algunos temas relacionados, usando el punto de vista de la continuación, una noción central en la Teoría del índice de Conley. Intuitivamente, decimos que cierta propiedad *continúa* si siempre que una familia parametrizada de flujos $\varphi_\lambda, \lambda \in I$, cumple que φ_0 tiene esta propiedad entonces φ_λ también la tiene para valores pequeños de λ. En esta sección, básicamente veremos que todo flujo uniformemente persistente tiene *propiedades de continuación débil*, en el sentido de que pequeñas perturbaciones del flujo nunca llevan a la extinción de poblaciones dentro de cierto rango (que puede ser arbitrariamente escogido). Además, si añadimos una condición de regularidad, que definiremos a continuación, sobre el flujo original se obtiene continuación completa.

Definición 1. Sea X espacio métrico localmente compacto y sea E un subconjunto cerrado de X. Sea φ_λ, $\lambda \in I$, una familia (continua) parametrizada de sistemas dinámicos disipativos en E para los que ∂E es invariante. Decimos que la familia es regular en λ_0 si existen un conjunto compacto $K \subset \mathring{E}$, $K \neq \emptyset$, y $\epsilon > 0$ tal que para todo $x \in \mathring{E}$ y para todo $\lambda \in (\lambda_0 - \epsilon, \lambda_0 + \epsilon)$, la trayectoria $\varphi_\lambda(x, \cdot)$ visita K.

A continuación enunciamos nuestros resultados de continuación, primero en una forma general y después consideramos el caso particular de flujos definidos en el ortante no negativo.

Teorema 5 (Continuación débil de la persistencia uniforme)**.** *Sea X un espacio métrico localmente compacto y sea E un subconjunto cerrado de X. Sea φ_λ, $\lambda \in I$, una familia (continua) parametrizada de sistemas dinámicos disipativos en E para los que ∂E es invariante. Supongamos que φ_0 es uniformemente persistente. Entonces existe $\beta > 0$ tal que para todo conjunto compacto $K \subset \mathring{E}$ existe $\lambda_0 > 0$ tal que* $\liminf\{d(\varphi_\lambda(x,t), \partial E) \mid t \to \infty\} \geq \beta$ *para todo $\lambda \leq \lambda_0$ y para todo $x \in K$.*

Corolario 1. *Sea φ_λ, $\lambda \in I$, una familia (continua) parametrizada de flujos disipativos en el ortante no negativo $\mathbb{R}^n_+$. Supongamos que φ_0 es uniformemente persistente. Entonces existe $\alpha > 0$ tal que para todo $\epsilon > 0$ y todo $M > \epsilon$, existe $\lambda_0 > 0$ tal que:*

a) $\liminf\{d(\varphi_\lambda(x,t), \partial E) \mid t \to \infty\} > \alpha$ *para todo x con $d(x, \partial\mathbb{R}^n_+) \geq \epsilon$ y $||x|| \leq M$ y para todo $\lambda \leq \lambda_0$, y*

b) el conjunto $W_\lambda = \{x \in \mathbb{R}^n_+ \mid \liminf\{d(\varphi_\lambda(x,t), \partial E) \mid t \to \infty\} > \alpha\}$ es contractible y el conjunto $R_\lambda = \{x \in \mathbb{R}^n_+ \mid \liminf\{d(\varphi_\lambda(x,t), \partial E) \mid t \to \infty\} \leq \alpha\} \cup \{\infty\}$ tiene la forma de $\mathbb{S}^{n-1}$ para todo $\lambda \leq \lambda_0$.

Teorema 6 (Continuación de la persistencia uniforme para familias regulares de sistemas disipativos)**.** *Sea X un espacio métrico localmente compacto y sea E un subconjunto cerrado X. Sea φ_λ, $\lambda \in I$, una familia de sistemas dinámicos disipativos en E para los que ∂E es invariante. Supongamos que la familia es regular en $\lambda = 0$ y que φ_0 es uniformemente persistente. Entonces existe $\lambda_0 > 0$ tal que φ_λ es uniformemente persistente para todo $\lambda \leq \lambda_0$. Además la continuación es uniforme en parámetros, en el sentido de que existe $\beta > 0$ tal que* $\liminf\{d(\varphi_\lambda(x,t), \partial E) \mid t \to \infty\} \geq \beta$ *para todo $\lambda \leq \lambda_0$ y para todo $x \in \mathring{E}$.*

Corolario 2. *Sea φ_λ, $\lambda \in I$, una familia (continua) parametrizada de flujos disipativos en el ortante no negativo $\mathbb{R}^n_+$. Supongamos*

a) existen ϵ y M, con $0 < \epsilon < M$, y $\lambda_0 > 0$ tal que para todo x en el interior de $\mathbb{R}^n_+$ y para todo $\lambda \leq \lambda_0$ existe un $t \in \mathbb{R}$ con $d(\varphi_\lambda(x,t), \partial\mathbb{R}^n_+) \geq \epsilon$ y $||\varphi_\lambda(x,t)|| \leq M$, y *b)* φ_0 *es uniformemente persistente.*

Entonces φ_λ *es uniformemente persistente para todo* $\lambda \leq \lambda_0$.

Sería interesante estudiar las implicaciones de estos resultados en algunas situaciones particulares. El Teorema 5 sugiere que la permanencia no desaparece *completamente* de manera abrupta. Incluso si el sistema no continúa, la permanencia se mantiene cuando nos limitamos a poblaciones dentro de cierto rango. Como caso interesante, S. Cano-Casanova y J. López-Gómez demuestran en [5] (ver también [30]) que la permanencia de dos especies es posible bajo agresión mutua fuerte. En otras palabras, demuestran que si las tasas de nacimiento son suficientemente altas entonces las especies son permanentes, independientemente de la fuerza de la competición en las regiones donde ésta tiene lugar. En concreto determinan cómo de alta debe ser la tasa de nacimientos. Un problema interesante sería el estudio de hasta qué punto la permanencia se conserva para poblaciones dentro de cierto rango a pesar de que sus tasas de reproducción estén por debajo del umbral límite.

4. Continuación uniforme de la persistencia uniforme

Discutiremos en esta sección algunos asuntos relacionados con la robustez, de nuevo usando el punto de vista de la continuación, en este caso usando una nueva noción de continuación que denominamos continuación uniforme.

Comenzamos introduciendo algunas nociones dinámicas que juegan un importante papel en el estudio de las propiedades relacionadas con la robustez de la persistencia uniforme.

Definición 2. Decimos que un compacto $M \subset \partial E$ es externamente repulsivo para un flujo $\varphi : E \times \mathbb{R} \longrightarrow E$ si existe un entorno U de M en E tal que para todo $x \in U - \partial E$ existe un $t > 0$ con $\varphi(x,t) \notin U$.

Si $\varphi_\lambda : E \times \mathbb{R} \longrightarrow E$ con $\lambda \in I$ es una familia parametrizada de flujos, decimos que $M \subset \partial E$ es uniformemente externamente repulsivo si existen un entorno U de M en E y $\lambda_0 > 0$ tal que para todo $x \in U - \partial E$ y todo $\lambda \in [0, \lambda_0]$ existe $t > 0$ con $\varphi_\lambda(x,t) \notin U$.

Definición 3. Sea $\varphi_\lambda : E \times \mathbb{R} \longrightarrow E$, $\lambda \in I$, una familia parametrizada de flujos. Supongamos que $M \subset \partial E$ es un conjunto invariante aislado para φ_0. Decimos que M es fuertemente aislado si existen un entorno U de M en E, $\lambda_0 > 0$ y un conjunto compacto $K \subset \mathring{E}$, tal que para todo $x \in U - \partial E$ y todo $\lambda \in [0, \lambda_0]$ la trayectoria $\varphi_\lambda(x, \cdot)$ visita K para algún $t \in \mathbb{R}$ (no necesariamente positivo).

Introducimos ahora la noción de continuación uniforme.

Definición 4. Sea $\varphi_\lambda : E \times \mathbb{R} \longrightarrow E$, $\lambda \in I$, una familia parametrizada de flujos. Decimos que la persistencia uniforme en $\lambda = 0$ continúa uniformemente si existen $\lambda_0 > 0$ y $\beta > 0$ tales que $\liminf_{t\to\infty} d(\varphi_\lambda(x,t), \partial E) \geq \beta$, para todo $x \in \mathring{E}$ y todo $\lambda \in [0, \lambda_0]$.

Es fácil ver que, en un contexto general, la persistencia uniforme no es una propiedad robusta. El ejemplo de la Figura 1 muestra que pequeñas perturbaciones de un flujo uniformemente persistente pueden destruir esta propiedad.

Esa familia de flujos satisface:

1) Es uniformemente disipativa.
2) φ_0 es uniformemente persistente, pero la persistencia uniforme en $\lambda = 0$ no continúa uniformemente. De hecho, para $\lambda > 0$, φ_λ no es uniformemente persistente.
3) El conjunto compacto invariante maximal M de φ_λ en ∂E es precisamente el punto fijo inferior; M es externamente repulsivo para φ_0, pero no es ni uniformemente externamente repulsivo ni fuertemente aislado.

El siguiente resultado proporciona condiciones necesarias y suficientes para la continuación uniforme de la persistencia uniforme.

Teorema 7. *Sea $\varphi_\lambda : E \times \mathbb{R} \longrightarrow E$, $\lambda \in I$, una familia (continua) parametrizada de flujos uniformemente disipativos. Supongamos que φ_0 es uniformemente persistente, y sea $\mathcal{M} = \{M_1, M_2, \ldots, M_n\}$ una descomposición de Morse para $\varphi_0|_M$ donde M es el conjunto compacto invariante maximal de φ_0 en ∂E. Entonces son equivalentes:*

a) La persistencia uniforme en $\lambda = 0$ continúa uniformemente.

b) $M_1, M_2, \ldots, M_n$ son fuertemente aislados.

Como consecuencia del Teorema 7 obtenemos la siguiente condición necesaria y suficiente para la continuación de la persistencia uniforme. Es una condición local y, por tanto, más fácil de verificar que el aislamiento fuerte, que es un concepto global.

Corolario 3. *Sea $\varphi_\lambda : E \times \mathbb{R} \longrightarrow E$, $\lambda \in I$, una familia (continua) parametrizada de flujos uniformemente disipativos. Supongamos que φ_0 es uniformemente persistente, y sea $\mathcal{M} = \{M_1, M_2, \ldots, M_n\}$ una descomposición de Morse para $\varphi_0|_M$ donde M es el conjunto compacto invariante maximal de φ_0 en ∂E. Entonces son equivalentes:*

a) La persistencia uniforme en $\lambda = 0$ continúa uniformemente.

b) $M_1, \ldots, M_n$ son uniformemente externamente repulsivos.

Observación 1. El Teorema 7 mejora el Teorema 6 de la sección anterior, pues sólo usamos información sobre descomposiciones de Morse del conjunto maximal en la frontera, mientras que entonces necesitábamos información sobre todos los puntos de $\mathring{E}$ (pedíamos que visitaran el conjunto compacto $K \subset \mathring{E}$).

En el Teorema 6 de la sección anterior sólo era necesaria la disipatividad para φ_λ.

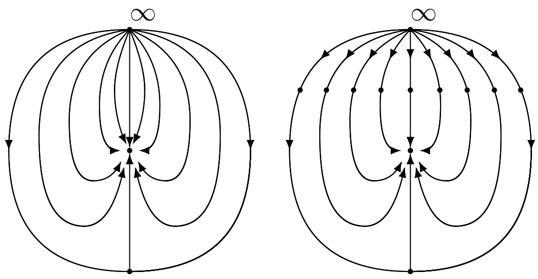


Figura 2: Un flujo uniformemente persistente φ_0 y una pequeña perturbación suya

La ilustración anterior muestra que la condición de disipatividad uniforme es esencial para la versión de este Teorema en términos de descomposiciones de Morse. La figura de la izquierda muestra el flujo para $\lambda = 0$. La figura de la derecha muestra el flujo para $\lambda > 0$. Este último tiene una línea de puntos fijos que se acerca al punto de infinito cuando λ se acerca a 0. Esta familia de flujos satisface:

1) Aunque todos los flujos son disipativos, la familia no es uniformemente disipativa.
2) φ_0 es uniformemente persistente pero la persistencia uniforme en $\lambda = 0$ no continúa uniformemente. De hecho, para $\lambda > 0$, φ_λ no es uniformemente persistente.
3) El conjunto compacto invariante maximal M de φ_0 en ∂E es precisamente el punto fijo inferior, y M es uniformemente externamente repulsivo y fuertemente aislado.

Podemos preguntarnos hasta qué punto las condiciones de las definiciones de compacto externamente repulsivo (resp. compacto uniformemente externamente repulsivo) pueden ser relajadas. Podríamos, en la Definición 2, pedir sólo que $\varphi(x,t) \notin U$ (resp. $\varphi_\lambda(x,t) \notin U$) para algún $t \in \mathbb{R}$ (*no necesariamente positivo*), i.e. que $U - \partial E$ no contenga trayectorias enteras de los flujos φ_λ. Esto daría lugar a una propiedad de aislamiento más débil que la de aislamiento fuerte de la Definición 3, que pide, además, que la trayectoria $\varphi_\lambda(x,\cdot)$ visite el compacto K. Sin embargo, si pedimos sólo esta condición más débil, entonces el Teorema 7 y el Corolario 3 dejan de ser ciertos. Esto es ilustrado de nuevo por el ejemplo de la Figura 1. A pesar de esto, aún tenemos cierta forma de continuación como muestra el siguiente resultado.

Denotamos por $\gamma_\lambda(x)$ la trayectoria de x para el flujo φ_λ.

Teorema 8. *Sea $\varphi_\lambda : E \times \mathbb{R} \to E$, $\lambda \in I$, una familia parametrizada de flujos uniformemente disipativa, donde φ_0 es uniformemente persistente. Sea M el conjunto compacto invariante maximal de $\varphi_0|_{\partial E}$. Supongamos que existen un entorno U de M en E y $\lambda_0 \in I$ tal que para todo $x \in U - \partial E$ y todo $\lambda \leq \lambda_0$ existe $t \in \mathbb{R}$ (no necesariamente positivo) tal que $\varphi_\lambda(x,t) \notin U$. Entonces la persistencia uniforme continúa uniformemente para órbitas acotadas en el sentido siguiente: existen $\lambda_1 \in I$ y $\beta > 0$ tales que* $\liminf_{t\to\infty}(d(\varphi_\lambda(x,t),\partial E)) > \beta$ *para todo $\lambda \leq \lambda_1$ y todo $x \in \mathring{E}$ tales que $\gamma_\lambda(x)$ es acotado.*

Es interesante destacar que la versión del Teorema 8 para descomposiciones de Morse de M no se cumple, como ilustra el siguiente ejemplo.

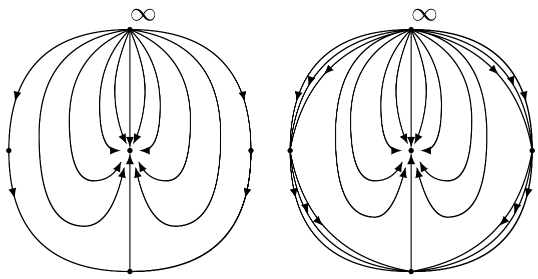


Figura 3: Un flujo uniformemente persistente φ_0 y una pequeña perturbación suya

La figura de la izquierda muestra el flujo para $\lambda = 0$. La figura de la derecha muestra el flujo para $\lambda > 0$. Para $\lambda = 0$ todas las órbitas son atraídas por el punto en el centro, con la excepción de tres puntos fijos en la frontera y las cuatro órbitas exteriores, atraídas, cada una, por uno de esos tres puntos. Para $\lambda > 0$ el comportamiento en ∂E es el mismo que para $\lambda = 0$ pero hay órbitas de puntos en $\mathring{E}$ que son atraídas, o repelidas, por los tres puntos fijos en la frontera. El conjunto de esas órbitas encoge según λ se acerca a 0. Esta familia de flujos satisface:

1) La familia de flujos es uniformemente disipativa.
2) Si M es el conjunto compacto invariante maximal de φ_0 en ∂E, existe una descomposición de Morse $\mathcal{M} = \{M_1, M_2, M_3\}$ para $\varphi_0|_M$, existe un entorno U de $M_1 \cup M_2 \cup M_3$ en E y existe $\lambda_0 \in I$, tal que para todo $x \in U - \partial E$ y todo $\lambda \leq \lambda_0$, existe $t \in \mathbb{R}$ (no necesariamente positivo) tal que $\varphi_\lambda(x,t) \notin U$.
3) φ_0 es uniformemente persistente pero la persistencia uniforme en $\lambda = 0$ no continúa uniformemente para órbitas acotadas en el sentido del Teorema 8.

Como consecuencia del Teorema 8 obtenemos el siguiente Corolario. Denotamos por $W_\lambda^-(U)$ el conjunto de todos los puntos $x \in E$ cuyo α-límite para el flujo φ_λ es no vacío y está contenido en U. En el caso particular en que U es el entorno de M en el enunciado del Teorema 8 entonces, para λ pequeño, $W_\lambda^-(U) = W_\lambda^-(M^\lambda)$, donde M^λ es la continuación de M para el flujo φ_λ.

Corolario 4. *Con las mismas hipótesis del Teorema 8, supongamos además que se cumple la siguiente condición: si $x \in \mathring{E}$ y $x \in W_0^-(U)$ entonces $x \in W_\lambda^-(U)$ para $\lambda \leq \lambda_0$. Entonces la persistencia uniforme continúa uniformemente para órbitas acotadas de φ_0, i.e., existen $\beta > 0$ y $\lambda_1 \in I$ tales que $\liminf_{t\to\infty} d(\varphi_\lambda(x,t), \partial E)) > \beta$ para todo $\lambda \leq \lambda_1$ y todo $x \in \mathring{E}$ tal que $\gamma_0(x)$ es acotado.*

5. Bifurcaciones generalizadas de Poincaré-Andronov-Hopf en puntos de la frontera de una n-variedad

Seibert y Seibert & Florio, a partir de trabajos anteriores de Marchetti, Negrini, Salvadori y Scalia, estudiaron en una serie de artículos las bifurcaciones de sistemas dinámicos resultantes de un cambio en la estabilidad de un punto fijo (ver [31, 49, 50, 51]). En particular estudiaron aquellas bifurcaciones que son una consecuencia de la transición de estabilidad asintótica a inestabilidad completa (sin necesidad de que las órbitas que se bifurcan sean periódicas). Supongamos que $\varphi_\lambda : E \times \mathbb{R} \to E$, $\lambda \in [0,1]$, es una familia parametrizada de flujos y $p \in E$ es un punto fijo de φ_λ, para todo $\lambda \in I$. Si p es un atractor para $\lambda = 0$ y un repulsor para $\lambda > 0$ entonces, de acuerdo con Seibert y Florio [50], una bifurcación de Poincaré-Andronov-Hopf generalizada tiene lugar en p.

Veremos que estas bifurcaciones juegan un papel relevante en la teoría de flujos uniformemente persistentes. El siguiente resultado describe las propiedades topológicas de las bifurcaciones generalizadas de Poincaré-Andronov-Hopf que ocurren en puntos de la frontera de $\mathbb{R}^n_+$ (o, más generalmente, de variedades con frontera). Ver

[47] para más propiedades de este tipo de bifurcaciones y [35] como referencia general sobre la bifurcación de Hopf.

Teorema 9. *Sea E una n-variedad conexa con frontera. Consideremos una familia continua de flujos $\varphi_\lambda : E \times \mathbb{R} \longrightarrow E$, $\lambda \in I$, y sea $p \in \partial E$ un punto fijo de φ_λ para todo $\lambda \in I$. Supongamos que p es un atractor para $\lambda = 0$ y un repulsor para $\lambda > 0$. Entonces para todo entorno compacto V de p contenido en la cuenca de atracción de p existe $\lambda_0 > 0$ tal que:*

a) Para todo λ, con $0 < \lambda \leq \lambda_0$, existe un atractor K_λ para φ_λ contenido en $\mathring{V} - \{p\}$. Además K_λ atrae todos los puntos de $V - \{p\}$.

b) K_λ tiene forma trivial.

c) El compacto $S_\lambda = K_\lambda \cap \partial E$ tiene la forma de $\mathbb{S}^{n-2}$ y K_λ / S_λ tiene la forma de $\mathbb{S}^{n-1}$.

d) K_λ descompone E en dos componentes conexas, $\mathcal{C}_\lambda$ y $\mathcal{D}_\lambda$, con $p \in \mathcal{C}_\lambda$ y tal que $\mathcal{C}_\lambda - \{p\}$ está contenido en la cuenca de atracción de K_λ.

e) La función multivaluada $\Phi : [0, \lambda_0] \longrightarrow V$ definida por $\Phi(0) = \{p\}$, $\Phi(\lambda) = K_\lambda$ (cuando $\lambda \neq 0$) es semicontinua superiormente.

Una observación importante, que se deduce de la demostración del Teorema anterior, es que $S_\lambda = K_\lambda \cap \partial E$ es, de hecho, un atractor para el flujo restricción $\varphi_\lambda|_{\partial E}$. Lo llamaremos la *corona* de la bifurcación de Poincaré-Andronov-Hopf para el valor del parámetro λ. Resulta que las propiedades de permanencia del flujo después de la bifurcación dependen fuertemente de esta corona. Nos referiremos al atractor K_λ que aparece en la bifurcación como el *Atractor de Hopf*.

Vemos a continuación un ejemplo de este tipo de bifurcaciones. Se construye combinando una interacción de tipo Holling con una bifurcación horca. Consiste en una familia parametrizada de flujos $\varphi_\lambda : \mathbb{R}^3_+ \times \mathbb{R} \longrightarrow \mathbb{R}^3_+$, donde φ_0 presenta un caso extremo de no permanencia: todo el interior es atraído por un punto en la frontera. Ejemplos de este tipo son habituales en dinámica de poblaciones.

Consideramos el sistema modelado por las ecuaciones:

$$\dot{x} = rx\left(1 - \frac{x}{K}\right) - y\frac{cx}{a+x} \qquad \dot{y} = y\left(-d + \frac{bx}{a+x}\right) \qquad \dot{z} = \frac{25(K-5)z - z^3}{s}$$

en $\mathbb{R}^3_+$, donde todos los parámetros son positivos. En particular, fijamos los valores de $a = 3$, $b = 4$, $c = 4$, $d = 1$, $r = 2$, $s = 4 \cdot 10^3$, y estudiamos el comportamiento del sistema en función del valor de K.

De acuerdo con [23], como $b > d$, siempre que $K > \dfrac{ad}{b-d} = 1$, restringido al plano xy, el sistema admite un único punto fijo interior (x_0, y_0) con $x_0 = \dfrac{ad}{b-d} = 1$. Además, (x_0, y_0) es un atractor global para el flujo restringido al cuadrante xy abierto si y sólo si $K \leq a + 2x_0 = 5$. Por otro lado, el flujo restringido al eje z tiene $z = 0$ como un atractor global siempre que $K \leq 5$. Por tanto, si $K \leq 5$, el punto $(x, y, 0)$ es un atractor global para el flujo restringido a $\{(x, y, z) \in \mathbb{R}^3_+ \mid x, y > 0\}$.

En particular, para $K = 5$, se tiene la situación ilustrada en la Figura 4.

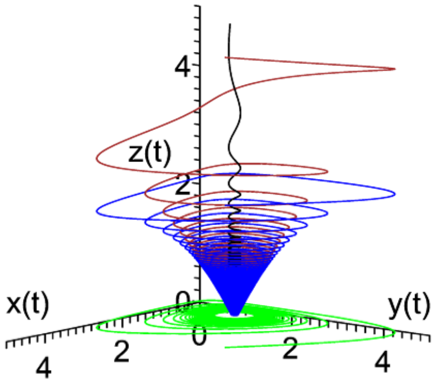


Figura 4: El flujo φ_K para $K = 5$

Para $K = 5$, el flujo restringido al plano xy experimenta una bifurcación de Hopf, (x_0, y_0) se convierte en repulsivo y un ciclo límite aparece alrededor de él. Por otro lado, el flujo restringido al eje z experimenta una bifurcación horca, $z = 0$ se convierte en repulsivo y aparece un punto fijo estable cuya cuenca de atracción es todo el conjunto $z > 0$. Por tanto, el flujo 3-dimensional completo experimenta una bifurcación generalizada de Poincaré-Andronov-Hopf: el punto $(x_0, y_0, 0)$ se convierte en repulsor, mientras que aparece un atractor, para el flujo completo en $\mathbb{R}^3_+$.

En la siguiente figura mostramos el comportamiento del flujo para algunos valores de $K > 5$.

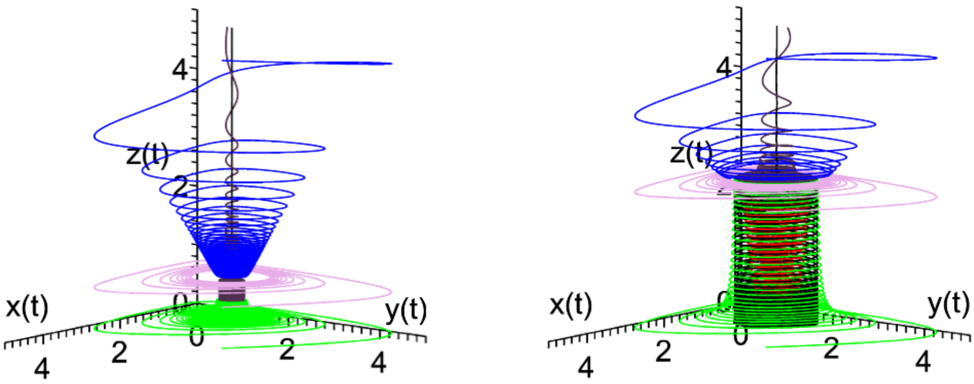


Figura 5: El flujo φ_K para $K = 5{,}02$ y $K = 5{,}2$

El atractor que aparece para $K > 5$ es la superficie de un cilindro sin la base inferior abierta (ver Figura 6). Es el atractor de Hopf definido de acuerdo con la demostración del Teorema 9. Su región de atracción es $\{(x, y, z) \in \mathbb{R}^3_+ \mid x, y > 0\}$ excepto el punto fijo inestable en el plano xy.

Por otro lado, la frontera de la base inferior es la corona, que es un atractor para el flujo restringido a la frontera de $\mathbb{R}^3_+$. Su región de atracción es el conjunto $\{(x, y, 0) \in \mathbb{R}^3_+ \mid x, y > 0\}$.

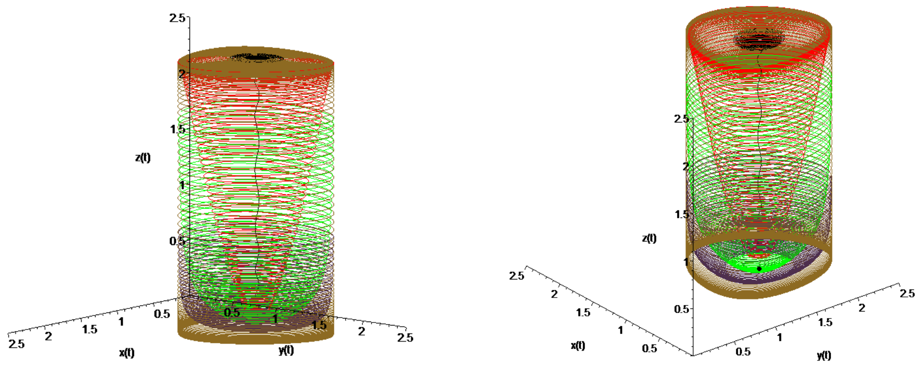


Figura 6: El atractor de Hopf y la corona para $K = 5{,}2$

Observamos en este ejemplo que después de que la bifurcación de Hopf tiene lugar, el flujo se transforma en un flujo uniformemente persistente: la parte superior del cilindro atrae todo los puntos del interior y es un atractor global interno. Esto ilustra un hecho general: la persistencia uniforme puede ser alcanzada en bifurcaciones de Hopf.

Nótese que la familia de flujos en el ejemplo no es uniformemente disipativa. Sin embargo, si excluimos los planos coordenados xz e yz, i.e. si consideramos los flujos definidos sólo en la variedad $\{(x, y, z) \in \mathbb{R}^3_+ \mid x, y > 0\}$ (que es invariante para todo los flujos) entonces la familia se convierte en uniformemente disipativa. Además, el hecho de que φ_λ sea uniformemente persistente para $\lambda > 0$ depende sólo del comportamiento de φ_λ en esta variedad invariante. De esta forma, el ejemplo es un caso particular simple del Teorema 10.

Consideramos ahora el caso general. Damos condiciones bajo las cuales un flujo con propiedades extremas de no permanencia se convierte en uniformemente persistente después de que una bifurcación de Poincaré-Andronov-Hopf tiene lugar.

Teorema 10. *Sea $\varphi_\lambda : E \times \mathbb{R} \to E$ una familia uniformemente disipativa de flujos definidos en la variedad n-dimensional con frontera, E, y sea M el conjunto compacto invariante maximal para φ_0. Supongamos que $M \subset \partial E$ (i.e. el atractor global está en la frontera) y sea $\mathcal{M} = \{M_1, M_2, \ldots, M_n\}$ una descomposición de Morse de M, donde $M_1 = \{p\}$. Supongamos que una bifurcación generalizada de Poincaré-Andronov-Hopf tiene lugar en p y que las siguientes condiciones se cumplen:*

a) $M_2, \ldots, M_n$ son uniformemente externamente repulsivos, y

b) la corona S_λ es externamente repulsiva para todo φ_λ con $0 < \lambda \leq \lambda_0$.

Entonces φ_λ es uniformemente persistente para $0 < \lambda \leq \lambda_0$ y el atractor de Hopf, K_λ, tiene una descomposición atractor-repulsor (A_λ, S_λ), donde $A_\lambda \subset \mathring{E}$ es el atractor *global interno y S_λ es la corona de la bifurcación. Recíprocamente, las condiciones a) y b) son necesarias para la persistencia uniforme de φ_λ para $0 < \lambda \leq \lambda_0$.*

6. El atractor global

Estamos interesados ahora en discutir algunas características de la dinámica dentro del atractor global de un flujo uniformemente persistente. Como motivación consideramos una interacción de tipo Holling, asociada a varios fenómenos en ecología y cinética química (ver [23]). Se modela por las ecuaciones

$$\dot{x} = rx\left(1 - \frac{x}{k}\right) - y\frac{cx}{a+x} \qquad \dot{y} = y\left(-d + \frac{bx}{a+x}\right)$$

en $\mathbb{R}^2_+$, donde todos los parámetros son positivos.

Si $b > d$ y $k > \dfrac{ad}{b-d}$, entonces el sistema admite un único punto fijo interior $F = (\bar{x}, \bar{y})$ con $\bar{x} = \dfrac{ad}{b-d}$.

Este punto fijo es un atractor global si y sólo si $k \leq a + 2\bar{x}$. Si $k > a + 2\bar{x}$ entonces F se convierte en un repulsor y una bifurcación de Hopf tiene lugar dentro del atractor global K y se crea un ciclo atractivo C que empieza su evolución desde F. Por tanto, el punto repulsivo F y el ciclo atractivo C inducen una descomposición atractor-repulsor del atractor global K. Nos gustaría estudiar este tipo de descomposiciones en una forma general. Esta situación tiene un rico significado topológico y la descripción de las propiedades de dualidad implicadas requiere el uso de la teoría de la forma. Sólo a este nivel los resultados se pueden expresar satisfactoriamente. En general, los resultados no son ciertos si reemplazamos forma por tipo topológico o incluso tipo de homotopía.

Introducimos primero algunas nociones que serán usadas posteriormente. Decimos que un continuo K es de tipo puntual en $\mathbb{R}^n$ si $\mathbb{R}^n - K$ es homeomorfo a $\mathbb{R}^n - \{p\}$, donde p es un punto. Un ejemplo importante de continuo de tipo puntual viene dado por los atractores globales en espacios euclídeos (ver [1]). Es bien sabido que un continuo de tipo puntual tiene forma trivial (i.e. la forma de un punto), aunque la implicación recíproca no es cierta en general.

Teorema 11. *Sea $\varphi : \mathbb{R}^n_+ \times \mathbb{R} \to \mathbb{R}^n_+$ un flujo disipativo. Si φ es uniformemente persistente entonces:*

a) Supongamos que L es un repulsor de tipo puntual (en particular un punto repulsivo) en el interior de $\mathbb{R}^n_+$, entonces existe un atractor K_0 con la forma de S^{n-1} contenido en el atractor global K y cuya cuenca de atracción es $\operatorname{int}\mathbb{R}^n_+ - L$.

b) Supongamos que L es un repulsor con la forma de S^{n-1} en el interior de $\mathbb{R}^n_+$. Entonces L descompone $\operatorname{int}\mathbb{R}^n_+$ en dos componentes conexas. Además, si la componente acotada es simplemente conexa entonces existe un atractor con la forma de un punto contenido (junto con su cuenca de atracción) en el interior del atractor global K.

Observación 2. No sabemos si la hipótesis sobre la conexión simple es necesaria en la parte b) del Teorema 11. De hecho existen inmersiones salvajes de la esfera en el espacio euclídeo 3-dimensional de modo que la componente conexa acotada del complementario no es simplemente conexa (por ejemplo, algunas variantes de la esfera cornuda de Alexander). Sin embargo, Sánchez-Gabites demostró en [44] que no pueden ser atractores ni repulsores de flujos en $\mathbb{R}^3$. De hecho, los resultados de Sánchez-Gabites pueden ser usados para probar que si un repulsor en $\mathbb{R}^3$ (o $\operatorname{int}\mathbb{R}^3_+$) tiene la forma de S^2 entonces la componente acotada del complementario es una bola abierta topológica (y como consecuencia del argumento utilizado en la demostración del Teorema 11, la región de atracción del atractor dual es homeomorfa a $\mathbb{R}^3$ y el atractor es de tipo puntual). En el caso 2-dimensional la componente acotada del complementario de un continuo con la forma de $\mathbb{S}^1$ es siempre un disco abierto topológico (ver [36]) y, por tanto, el atractor dual de L es de tipo puntual. Por consiguiente, en el caso de dimensiones bajas la hipótesis de conexión simple es innecesaria.

Por otro lado, es fácil construir un flujo uniformemente persistente en $\mathbb{R}^2_+$ con un punto repulsivo y tal que el atractor complementario es el círculo polaco (ver Hastings [19]). Esto muestra que, en general, el Teorema 11 no puede ser mejorado para probar que el atractor es homeomorfo a S^{n-1}. De hecho, ni siquiera es posible demostrar que es homotópicamente equivalente a S^{n-1}.

En el siguiente resultado veremos que la teoría de Morse de flujos uniformemente persistentes con un ciclo atractivo es bastante directa, independientemente de la complejidad del flujo en la frontera.

Supongamos que $\varphi : \mathbb{R}^n_+ \times \mathbb{R} \to \mathbb{R}^n_+$ es un flujo uniformemente persistente. Decimos que $\mathcal{M} = \{M_1, M_2, \ldots, M_k\}$ es una descomposición *natural* de Morse del flujo si

1) $\{M_1, M_2\}$ es una descomposición atractor-repulsor del atractor global K,
2) $M_i \subset \partial\mathbb{R}^n_+$ para $i \geq 3$, y
3) $\{M_1, M_2, \ldots, M_k, \infty\}$ es una descomposición de Morse de $\mathbb{R}^n_+ \cup \{\infty\}$.

Por ecuación de Morse de $\mathcal{M}$ nos referimos a la de $\{M_1, M_2, \ldots, M_k, \infty\}$.

El siguiente teorema muestra que si M_1 es un ciclo atractivo o, más generalmente, un atractor con la forma de $\mathbb{S}^1$, entonces la ecuación de Morse de $\mathcal{M}$ toma una forma simple. En la dirección opuesta veremos que, usando esta ecuación, podemos reconocer la existencia de atractores con la forma de $\mathbb{S}^1$ en el plano, o atractores cuya suspensión tiene la forma de $\mathbb{S}^2$ para dimensiones más altas.

Teorema 12. *Sea $\varphi : \mathbb{R}^n_+ \times \mathbb{R} \to \mathbb{R}^n_+$ un flujo disipativo. Supongamos que φ es uniformemente persistente y que $\mathcal{M} = \{M_1, M_2, \ldots, M_k\}$ es una descomposición natural de Morse de $\mathbb{R}^n_+$ para φ. Entonces:*

a) Si M_1 tiene la forma de $\mathbb{S}^1$, entonces la ecuación de Morse de la descomposición $\mathcal{M}$ con coeficientes en $\mathbb{Z}$ o en un cuerpo es $1 + t + t^2 = 1 + (1+t)t$.

b) Recíprocamente, si la ecuación de Morse de $\mathcal{M}$ es la anterior, entonces $Sh(M_1) = Sh(\mathbb{S}^1)$, para $n = 2$, y $Sh(\Sigma M_1) = Sh(\mathbb{S}^2)$, para $n \geq 2$, donde ΣM_1 es la suspensión de M_1.

Un resultado análogo puede ser demostrado para dimensiones más altas.

Teorema 13. *Sea $\varphi : \mathbb{R}^n_+ \times \mathbb{R} \to \mathbb{R}^n_+$ un flujo disipativo. Supongamos que φ es uniformemente persistente y que $\mathcal{M} = \{M_1, M_2, \ldots, M_k\}$ es una descomposición natural de Morse de $\mathbb{R}^n_+$ para φ. Entonces:*

a) Si M_1 tiene la forma de $\mathbb{S}^r$ entonces la ecuación de Morse de la descomposición $\mathcal{M}$ con coeficientes en $\mathbb{Z}$ o en un cuerpo es $1 + t^r + t^{r+1} = 1 + (1+t)t^r$.

b) Recíprocamente, si la ecuación de Morse de $\mathcal{M}$ es la anterior, entonces $Sh(\Sigma M_1) = Sh(S^{r+1})$, para $n \geq 1$, donde ΣM_1 es la suspensión de M_1.

Los teoremas 12 y 13 no se cumplen si reemplazamos la suspensión ΣM_1 del atractor por el atractor M_1. Es fácil definir un flujo uniformemente persistente en $\mathbb{R}^3_+$ donde el atractor global K es una bola con una descomposición atractor-repulsor tal que el repulsor R es un arco anudado con extremos en ∂K y el atractor dual A es un ANR. Entonces A tiene la homología de $\mathbb{S}^1$, pero grupo fundamental no isomorfo a $\mathbb{Z}$. Por tanto $Sh(A) \neq Sh(\mathbb{S}^1)$, a pesar del hecho que la ecuación de Morse de cualquier descomposición natural de Morse con $M_1 = A$, $M_2 = R$ es $1 + t + t^2 = 1 + (1+t)t$.

En su artículo [27] D. Li ha desarrollado nuevas ideas para el estudio de descomposiciones de Morse desde el punto de vista de las funciones de Morse-Lyapunov. Sería interesante aplicar las ideas de Li al contexto actual.

Referencias

[1] N. P. Bhatia, G. P. Szego: *Stability theory of dynamical systems.* Grundlehren der Mat. Wiss. 16, Springer, Berlin, 1970.

[2] K. Borsuk: *Theory of Shape.* Monografie Matematyczne 59, Polish Scientific Publishers, Warsaw, 1975.

[3] G. Butler, H. I. Freedman, P. Waltman: *Uniformly persistent systems.* Proc. AMS 96 (1986) 425–430.

[4] G. Butler and P. Waltman: *Persistence in dynamical systems.* J. Differential Equations 63 (1986) 255–263.

[5] S. Cano-Casanova and J. López-Gómez: *Permanence under strong aggressions is possible.* Ann. Inst. H. Poincaré Anal. Non Linéaire 20 (2003) 999–1041.

[6] C. Conley: *Isolated invariant sets and the Morse index.* CBMS Regional Conference Series in Mathematics 38, AMS Providence, RI.

[7] C. Conley, E. Zehnder: *Morse-type index theory for flows and periodic solutions for Hamiltonian equations.* Comm. Pure Appl. Math. 37 (1984) 207–253.

[8] J. M. Cordier, T. Porter: *Shape theory. Categorical methods of approximation.* Mathematics and its Applications, Ellis Horwood Ltd., Chichester; Halsted Press [John Wiley & Sons, Inc.], New York 1989.

[9] A. Dold: *Lectures on Algebraic Topology.* Springer-Verlag 1995.

[10] J. Dydak, J. Segal: *Shape theory. An introduction.* LNM 688, Springer-Verlag, Berlin 1978.

[11] B. M. Garay: *Uniform persistence and chain recurrence.* J. Math. Analysis Appl. 139 (1989) 372–381.

[12] B. M. Garay, J. Hofbauer: *Robust permanence for ecological differential equations, minimax, and discretizations.* SIAM J. Math. Anal. 34 (2003) 1007–1039.

[13] A. Giraldo, V. F. Laguna, J. M. R. Sanjurjo: *Uniform persistence and Hopf bifurcations in* R^n_+. J. Differential Equations 256 (8) (2014) 2949-2964.

[14] A. Giraldo, M. Morón, F. R. Ruiz del Portal, J. M. R. Sanjurjo: *Shape of global attractors in topological spaces.* Nonlinear Anal. 60, 5 (2005) 837–847

[15] A. Giraldo, J. M. R. Sanjurjo: *On the global structure of invariant regions of flows with asymptotically stable attractors.* Math. Z. 232, 4 (1999) 739–746.

[16] A. Giraldo, J. M. R. Sanjurjo: *Singular continuations of attractors.* SIAM J. Appl. Dyn. Syst. 8 (2009) 554–575.

[17] B. Günther, J. Segal: *Every attractor of a flow on a manifold has the shape of a finite polyhedron.* Proc. AMS 119 (1993) 321–329.

[18] J. K. Hale: *Stability and gradient dynamical systems.* Rev. Mat. Complut. 17 (2004) 7–57.

[19] H. M. Hastings: *A higher-dimensional Poincaré-Bendixson theorem.* Glas. Mat. Ser. III 14 (34) no. 2 (1979) 263–268.

[20] A. Hatcher: *Algebraic topology.* Cambridge University Press, Cambridge, 2002.

[21] M. W. Hirsch, H. L. Smith, X-Q. Zhao: *Chain transitivity, attractivity, and strong repellors for semidynamical systems.* J. Dynam. Differential Equations 13 (2001) 107–131.

[22] J. Hofbauer: *A unified approach to persistence.* Acta Applicandae Mathematicae 14 (1989) 11–22.

[23] J. Hofbauer, K. Sigmund: *The theory of evolution and dynamical systems. Mathematical aspects of selection.* London Mathematical Society Student Texts, 7, Cambridge University Press, Cambridge, 1988.

[24] S. T. Hu: *Theory of retracts.* Wayne State Universite Press, Detroit, 1965.

[25] V. Hutson: *The stability under perturbations of repulsive sets.* J. Differential Equations 76 (1988) 77–90.

[26] L. Kapitanski, I. Rodnianski: *Shape and Morse theory of attractors.* Comm. Pure Appl. Math. 53 (2000) 218–242.

[27] D. Li: *A functional approach towards Morse theory of attractors for infinite dimensional dynamical systems.* Preprint.

[28] D. Li: *Morse Theory of Attractors for Infinite Dimensional Dynamical Systems via Lyapunov Functions.* Preprint.

[29] D. Li, A. Qi: *Morse equation of attractors for nonsmooth dynamical systems.* J. Differential Equations 253 (11) (2012) 3081–3100.

[30] J. López Gómez: *Strong competition with refuges.* Nonlinear Anal. 30 (1997) 5167–5178.

[31] F. Marchetti, P. Negrini, L. Salvadori, M. Scalia: *Liapunov direct method in approaching bifurcation problems*. Ann. Mat. Pura Appl. (4) 108 (1976) 211–226.

[32] S. Mardesic, J. Segal:*Shape theory. The inverse system approach*. North-Holland Mathematical Library, 26, North-Holland Publishing Co., Amsterdam-New York 1982.

[33] S. Mardesic: *Strong shape and homology*. Springer Monographs in Mathematics, Springer-Verlag, Berlin 2000.

[34] S. Mardesic: *Pairs of compacta and trivial shape*. Trans. AMS 189 (1971) 329–330.

[35] J. E. Marsden, M. McCracken: *The Hopf Bifurcation and Its Applications*. Springer, New York 1976.

[36] M. A. Morón, J. J. Sánchez-Gabites, J. M. R. Sanjurjo: *Topology and dynamics of unstable attractors*. Fund. Math. 197 (2007) 239–252.

[37] S. Nowak: *Algebraic theory of fundamental dimension*. Dissertationes Math. 187 (1981), 54 pp.

[38] J. W. Robbin, D. Salamon: *Dynamical systems, shape theory and the Conley index*. Ergodic Theory Dynam. Systems 8 (Charles Conley Memorial Volume) (1988) 375–393.

[39] J. C. Robinson: *Global attractors: topology and finite-dimensional dynamics*. J. Dynam. Differential Equations 11, 3 (1999) 557–581.

[40] J. C. Robinson: *Infinite-dimensional dynamical systems. An introduction to dissipative parabolic PDEs and the theory of global attractors*. Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge 2001.

[41] D. Salamon: *Connected simple systems and the Conley index of isolated invariant sets*. Trans. AMS 291 (1985) 1–41.

[42] J. J. Sánchez-Gabites: *Unstable attractors in manifolds*. Trans. AMS 362 (2010) 3563–3589.

[43] J. J. Sánchez-Gabites: *An approach to the shape Conley index without index pairs*. Rev. Mat. Complut. 24 (2011) 95–114.

[44] J. J. Sánchez-Gabites: *How strange can an attractor for a dynamical system in a 3-manifold look?* Nonlinear Analysis 74 (17) (2011) 6162–6185.

[45] J. M. R. Sanjurjo: *On the structure of uniform attractors*. J. Math. Anal. Appl. 192 (1995) 519–528.

[46] J. M. R Sanjurjo: *Morse equations and unstable manifolds of isolated invariant sets*. Nonlinearity 16 (2003) 1435–1448.

[47] J. M. R. Sanjurjo: *Global topological properties of the Hopf bifurcation*. J. Differential Equations 243 (2007) 238–255.

[48] J. M. R. Sanjurjo: *On the fine structure of the global attractor of a uniformly persistent flow*. J. Differential Equations 252 (2012) 4886–4897.

[49] P. Seibert: *Dynamical persistence principles and bifurcation*. Universitatis Iagellonicae Acta Mathematica XXXVI (1998) 65–77.

[50] P. Seibert, J. S. Florio: *On the foundations of bifurcation theory*. Nonlinear Anal. 22 (1994) 927–944.

[51] P. Seibert, J. S. Florio: *On Bifurcations Arising from Unstable Equilibria and Invariant Sets*. J. Differential Equations 150 (1998) 250–263.

[52] S. J. Schreiber: *Criteria for C^r-robust permanence*. J. Differential Equations 162 (2000) 400–426.

[53] H. L. Smith, H. R. Thieme: *Dynamical systems and population persistence*. Graduate Studies in Mathematics, AMS, Providence, RI, 2011.

[54] E. Spanier: *Algebraic Topology*. McGraw-Hill, New York 1966.

[55] S. Spiez: *On a characterization of shapes of several compacta*. Bull. Acad. Polon. Sci 21 (1973) 727–733.

[56] R. Temam: *Infinite Dimensional Dynamical Systems in Mechanics and Physics*. Applied Mathematical Sciences 68, Springer-Verlag, New York 1988.

[57] K. Wójcik: *Conley index and permanence in dynamical systems*. Topol. Methods Nonlinear Anal. 12 (1998) 153–158.

2-dimensional stratifolds

J. C. Gómez-Larrañaga,* **F. González-Acuña** and **Wolfgang Heil**

Centro de Investigación en Matemáticas
A.P. 402, Guanajuato 36000, Gto. México
jcarlos@cimat.mx

Instituto de Matemáticas UNAM
62210 Cuernavaca, Morelos, México
and Centro de Investigación en Matemáticas
A.P. 402, Guanajuato 36000, Gto. México
fico@math.unam.mx

Department of Mathematics
Florida State University
Tallahassee, FL 32306, USA
heil@math.fsu.edu

Dedicado a José María Montesinos Amilibia con motivo de su 70 aniversario.

ABSTRACT

2-stratifolds are a generalization of 2-manifolds in that there are disjoint simple closed curves where several sheets meet. They arise in the study of categorical invariants of 3-manifolds and may have applications to topological data analysis. We define 2-stratifolds and study properties of their associated labeled graphs that determine which ones are simply connected.

2010 Mathematics Subject Classification: 57M20, 57M05, 57M15
Key words: stratifold, simply connected, graph.

*Author supported by the Topological Data Analysis Fund of CIMAT.

1. Introduction

Topologically stratified spaces were introduced by René Thom [11], who showed that every Whitney stratified space was also a topologically stratified space, with the same strata. Later, Shmuel Weinberger [12] studied the topological classification of stratified spaces. In differential topology the concept of stratifolds, one of the many concepts of stratified spaces, was defined by Mathias Kreck [5]. In topological data analysis [C] one applies topological methods to study complex high dimensional data sets by extracting shapes (patterns) and obtaining insights about them. In these applications, some of these shapes turn out to be simple 2-dimensional simplicial complexes where it is computationally possible to calculate topological invariants such as the number of connected components or fundamental cycles. It happens that these invariants reflect valuable information from data. Some examples of this methodology to analyze data can be found in [7] and [2].

In this paper, we define closed topological 2-stratifolds and start their study having in mind that these objects could be good models for topological data analysis. A 2-stratifold X contains a collection of finitely many simple closed curves (the components of the 1-skeleton X_1 of X) such that $X - X_1$ is a 2-manifold and a neighborhood of each component C of X_1 consists of sheets (the precise definition is given in section 2). While studying categorical group invariants of 3-manifolds [4], these 2-stratifolds appear in a natural way. For instance, if $\mathcal{G}$ is a non-empty family of groups that is closed under subgroups, one would like to determine which (closed) 3-manifolds have $\mathcal{G}$-category equal to 3. It can be shown that such manifolds have a decomposition into three compact 3-submanifolds H_1, H_2, H_3 , where the intersection of $H_i \cap H_j$ (for $i \neq j$) is a compact 2-manifold, and each H_i is $\mathcal{G}$-contractile (i.e. for each connected component C of H_i the image of the fundamental group of C in the fundamental group of the 3-manifold is in the family $\mathcal{G}$). The nerve of this decomposition is the union of all the intersections $H_i \cap H_j$ $(i \neq j)$. This nerve is a 2-stratifold and plays an important role in studying categorical invariants. In contrast to 2-manifolds there is no known classification of 2-stratifolds in terms of their fundamental group. As a first step in this direction we ask which 2-stratifolds are 1-connected and we give a complete solution for the case when the associated graph (see below) is linear.

An interesting special class of 2-complexes, called *foams*, has been defined and studied by Khovanov [6]. These were also studied by Carter [3], and they are essentially 2-stratifolds for which a neighborhood of each point of the 1-skeleton consists of three sheets. We call these *trivalent* 2-stratifolds and develop an algorithm that detects whether such a given 2-stratifold is simply connected.

Here is an outline of the paper: In section 2 we define (closed) 2-stratifolds X and their associated bicolored graphs $G(X)$, and show that $G(X)$ is a retract of X. The graph $G(X)$ consists of "white" and "black" vertices that correspond to the 2-

manifold parts and the 1-skeleton, respectively. Together with a labeling of the edges, $G(X)$ essentially determines X. In section 3 we show that the graph of a 1-connected 2-stratifold is a tree such that all white vertices have genus 0 and all terminal vertices are white. Then it is shown that a 1-connected 2-stratifold is homotopy equivalent to a wedge of $m-n$ 2-spheres, where m and n are the number of white resp. black vertices of the tree $G(X)$. In section 4 we classify all 1-connected 2-stratifolds whose associated graph is linear. Finally, in section 5, we develop an algorithm to decide whether a trivalent 2-stratifold is simply connected.

2. 2-stratifolds and their Graphs

A (2-layer) 2-stratifold is a compact, connected Hausdorff space X together with a filtration $\emptyset = X_0 \subset X_1 \subset X_2 = X$ by a closed subspace such that X_1 is a closed 1-manifold, each point $x \in X_1$ has a neighborhood homeomorphic to $\mathbb{R}\times CL$, where CL is the open cone on L for some (finite) set L of cardinality > 2 and each $x \in X_2/X_1$ has a neighborhood homeomorphic to $\mathbb{R}^2$.

A component $C \approx S^1$ of X_1 has a regular neighborhood $N(C) = N_\pi(C) := (Y\times[0,1])/(y,1) \sim (h(y),0)$. Here Y is the closed cone on the discrete space $\{1,2,...,d\}$ and $h : Y \to Y$ is a homeomorphism whose restriction to $\{1,2,...,d\}$ is the permutation $\pi : \{1,2,...,d\} \to \{1,2,...,d\}$. If π' is conjugate to π in S_d, the group of permutations of $\{1,2,...,d\}$, then $N_\pi(C) = N_{\pi'}(C)$. So we may think of π as a conjugate class in S_d , that is, as a partition of d. A component of $\partial N_\pi(C)$ (the set of points having an open neighborhood homeomorphic to $R\times[0,1)$) corresponds to a summand of the partition π. It follows from Theorems 2 and 3 of [8] that $N(C) = N_\pi(C)$ for a unique partition π.

Another construction of $N_\pi(C)$ is as follows: If π is the partition $n_1 + n_2 + ... + n_r$ of d, let $f : \tilde{C} \to C$ be the covering of C where $\tilde{C}$ has components $\tilde{C}_1, \tilde{C}_2, \dots, \tilde{C}_r$ and the restriction of f to $\tilde{C}_i$ is an n_i-fold covering ($i = 1,...,r$). Then $N_\pi(C)$ is the mapping cylinder of f.

We take the neighborhoods of the components of X_1 sufficiently small so that $N(C) \cap N(C') = \emptyset$ if $C' \neq C$. Call the closures of the components of $N(C) - C$ the *sheets* of $N(C)$.

We now construct the connected bipartite graph $G(X, X_1)$ associated to the 2-stratifold (X, X_1).

Write $M = \overline{X - \cup_j N(C_j)}$ where C_j runs over the components of X_1. The white vertices of the graph $G(X, X_1)$ are the components of M; the black vertices are the $N(C_j)$'s . An edge is a component S of ∂M; it joins a white vertex W with a black vertex $N(C_j)$ if $S = W \cap N(C_j)$.

We obtain a geometric realization of $G(X, X_1)$ as an embedding into X as follows: In each component C_j of X_1 choose a black vertex b_j, in the interior of each component W_i of $M = \overline{X - \cup_j N(C_j)}$ choose a white vertex w_i. In each component S_{ij} of $W_i \cap N(C_j)$ choose a point y_{ij}, an arc α_{ij} in W_i from w_i to y_{ij} and an arc β_{ij} from y_{ij} to b_j in the sheet of $N(C_j)$ containing y_{ij}. An edge e_{ij} between w_i and b_j consists of the arc $\alpha_{ij} * \beta_{ij}$. For a fixed i, the arcs α_{ij} are chosen to meet only at w_i.

Lemma 1. *There is a retraction $r : X \to G(X, X_1)$.*

Proof. For each component W_i of M, the union of the arcs $\bigcup_j \alpha_{ij}$ is a cone on the points y_{ij} with cone point w_i, so $\bigcup_j \alpha_{ij}$ can be considered as a retract of $I \times I$ and therefore it has the Tietze extension property. Thus the map $\bigcup_j \alpha_{ij} \cup \partial W_i \to \bigcup_j \alpha_{ij}$ that is the identity on $\bigcup_j \alpha_{ij}$ and retracts ∂W_i to $\bigcup\{y_{ij}\}$ extends to a retraction of W_i to $\bigcup_j \alpha_{ij}$. Similarly there is a retraction of $N(C_j)$ to $\bigcup_i \beta_{ij}$. Since the retractions agree on $W_i \cap N(C_j) \to \{y_{ij}\}$ and combine to yield a retraction $r : X \to G(X, X_1)$. $\square$

We now assign labels to the edges of $G(X, X_1)$.

If F is a compact surface, $g(F)$ will be the genus, with the convention that $g(F) < 0$ if F is nonorientable. Thus $g(P^2) = g(Mb) = -1,\ g(K) = -2$, and so on, where Mb is the Moebius band and K is the Klein bottle. This convention follows Neumann [10].

A white vertex W is labeled with the genus g of W (as defined above). An edge S is labeled by n, where n is the summand of the partition π corresponding to the component S of $\partial N_\pi(C)$ where $S \subset \partial N_\pi(C)$.

Note that there is no need to label a black vertex with the partition π because the partition is shown by the labels of the adjacent edges. Also note that the the number of boundary components of W is the number of adjacent edges of W. If $G = G(X, X_1)$ is a tree, or if all white vertices have labels $g < 0$, then it is easy to see that the labeled graph determines X uniquely. If the graph contains vertices with non-negative genus one needs some additional structure (see [10]): let G^* be the subgraph of G defined by all non-terminal vertices with $g \geq 0$, that is, the union of all edges incident on such vertices and assign a $+$ or $-$ sign to each edge on $G^*(X, X_1)$. This information determines a cocyle $\epsilon_G \in H^1(G^*; Z/2)$ which assigns to any cycle the number modulo 2 of $(-)$-edges on that cycle. This is the additional structure on $G(X, X_1)$ needed to recover (X, X_1).

3. Simply connected 2-stratifolds

Let X be a 2-stratifold with associated bicolored graph $G = G(X, X_1)$.

By a subcomplex T of X we mean that $T = r^{-1}(\Gamma)$, where Γ is a subgraph of G and r is the retraction of Lemma 1. Let $\hat{T}$ be the quotient of X obtained by collapsing to a point the closure of each component of $X - T$. Note that $\hat{T}$ is a 2-stratifold whose graph $\hat{\Gamma}$ is is the union of Γ and the labeled edges (with their vertices) of $st(\Gamma) - \Gamma$ which are adjacent to a black vertex of Γ. (Here $st(\Gamma)$ is the star of Γ).

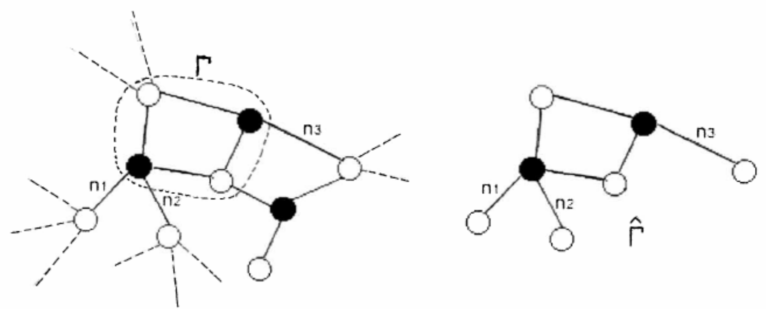


Figure 1: T and $\hat{T}$

Then $\pi(\hat{T})$ is a quotient of $\pi(X)$ and similarly $H_1(\hat{T})$ and $H_1(\hat{T}; \mathbb{Z}_2)$ are quotients of $H_1(X)$ and $H_1(X); \mathbb{Z}_2)$, respectively. In particular, if X is simply connected, so is $\hat{T}$.

Theorem 1. *(a) If $H_1(X)$ is finite, then G is a tree.*
(b) If $H_1(X, \mathbb{Z}_2) = 0$, then all white vertices of G have genus 0.
(c) If $H_1(X) = \mathbb{Z}_2^m$ for some $m \geq 0$, then all terminal vertices are white.

Corollary 1. *If X is simply connected, then G is a tree, all white vertices of G have genus 0, and all terminal vertices are white.*

Proof. (a) The retraction of Lemma 1 induces an epimorphism $r_* : H_1(X) \to H_1(G)$ and it follows that G is a tree.

(b) If T is a white vertex then $\hat{T}$ is a closed 2-manifold and since $H_1(X, \mathbb{Z}_2) = 0$, the genus associated to T is 0.

(c) Note that a terminal black vertex T of G corresponds to $\partial N_\pi(C)$ (with $\partial N_\pi(C)$ connected) and the label n of the edge adjacent to T is the order of π (an n-cycle) defining $N_\pi(C)$, hence $n \geq 3$ and $H_1(\hat{T}) \cong \mathbb{Z}_n$. By the hypothesis this does not occur. □

We have the following homotopy classification of 1-connected 2-stratifolds:

Theorem 2. *If X is simply connected then X is homotopy equivalent to a wedge of $m-n$ 2-spheres, where m (resp. n) is the number of white (resp. black) vertices of the tree G.*

In particular, two simply connected 2-stratifolds have the same homotopy type if and only if they have the same "deficiency" $n-m$ and no 2-stratifold is contractible.

Proof. To see that X is homotopy equivalent to a wedge of 2-spheres, we follow Milnor's proof of Theorem 6.5 in [9]: Since $\pi_1(X) = 1$ it follows from Hurewicz that $\pi_2(X) \cong H_2(X)$. Now $H_2(X)$ is free abelian (since torsion elements come from $H_3(X) = 0$). The maps $(S^2, *) \to (X, *)$ representing the generators of $\pi_2(X, *)$ combine to a map $f : S^2 \vee \cdots \vee S^2 \to X$ that induces an isomorphism $f_* : \pi_2(S^2 \vee \cdots \vee S^2) \to \pi_2(X)$. Now it follows from Whitehead that f is a homotopy equivalence.
To count the number of spheres in the wedge, recall that $X = M \cup N$, where $M = \overline{X - N}$ and $N = \cup_j N(C_j)$ (and C_j are the components of X_1). The Euler characteristic $\kappa(X) = \kappa(M) + \kappa(N) - \kappa(M \cap N) = \beta_0 - \beta_1 + \beta_2$, where β_i is the i's Betti number of X. Since $\kappa(N) = \kappa(M \cap N) = 0$ and $\beta_1 = 0$, it follows that $\beta_2 = \kappa(M) - 1$, which is also the number of 2-spheres in $S^2 \vee \cdots \vee S^2$. Now M is a (disjoint) union of m punctured spheres and the total number of punctures is the number e of edges of G. Therefore $\kappa(M) = 2m - e$. Since G is a tree, the number of vertices of G is $m + n = e + 1$. It follows that $\beta_2 = m - n$. □

4. Linear 2-stratifolds

We now consider the 2-stratifold $X(m_1, n_1, m_2, n_2, \ldots, m_r)$ whose associated graph $G = G(m_1, n_1, m_2, n_2, \ldots, m_r)$ is the linear graph with successive vertices $w_0, b_1, w_1, b_2, w_2, \ldots, b_r, w_r$ and successive labels $m_1, n_1, m_2, n_2, \ldots, m_r, n_r$ where m_i (resp. n_i) is the label of the edge joining b_i to w_{i-1} (resp. w_i) for $i = 1, \ldots, r$ and all white vertices $w_0, w_1, \ldots w_r$ have genus 0.

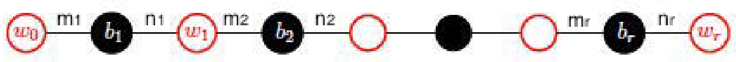


Figure 2: $G(m_1, n_1, m_2, n_2, \ldots, m_r)$

The fundamental group of X is the group

$$G_{m_1,n_1,\ldots,m_r,n_r} = \{x_1, \ldots, x_r : \; x_1^{m_1} = 1,\, x_1^{n_1} = x_2^{m_2}, \ldots, x_{r-1}^{n_{r-1}} = x_r^{m_r},\, x_r^{n_r} = 1\}$$

Lemma 2. *For integers m_i, n_j where $i = 1, \dots r$, $j = 1, \dots, r-1$, let*

$$G_r = \{x_1, \dots, x_r : \ x_1^{m_1} = 1, \, x_1^{n_1} = x_2^{m_2}, \dots, x_{r-1}^{n_{r-1}} = x_r^{m_r}\}.$$

If $gcd(m_i, n_j) = 1$ for $i \leq j \leq r-1$, then $G_r \cong \mathbb{Z}_m$, where $m = m_1 \cdots m_r$.

Proof. First we note that
$(*)$ $\qquad x_i^{m_1 m_2 \dots m_i} = 1$ for $i = 1, ..., r$
This is trivial for $i = 1$. By induction, $1 = x_{i-1}^{m_1 \dots m_{i-1}}$, and from $x_{i-1}^{n_{i-1}} = x_i^{m_i}$ it follows that $1 = x_{i-1}^{n_{i-1} m_1 \dots m_{i-1}} = x_i^{m_i m_1 \dots m_{i-1}}$.
Since for $i = 1, \dots, r-1$, $x_i^{n_i} \in \langle x_{i+1} \rangle$ (the cyclic subgroup generated by x_{i+1}) and $gcd(n_i, m_1 \dots m_i) = 1$, it follows from $(*)$ that $x_i \in \langle x_{i+1} \rangle$, hence G_r is cyclic generated by x_r. To see that the order of x_r is not less than $m_1 \cdots m_r$, we observe that the relation matrix

$$\begin{pmatrix} m_1 & & & & & \\ -n_1 & m_2 & & & & \\ & -n_2 & m_3 & & & \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & -n_{r-1} & m_r \end{pmatrix}$$

can be reduced to a diagonal matrix by elementary row and column operations, which leave the determinant invariant. Since G_r is cyclic, all diagonal entries of the normal form are $= 1$ except for the last entry, which is the order of G_r. $\square$

Corollary 2. *If $gcd(m_i, n_j) = 1$ for $i \leq j \leq r$, then the group*

$$G_{m_1, n_1, \dots, m_r, n_r} = \{x_1, \dots, x_r : \ x_1^{m_1} = 1, \, x_1^{n_1} = x_2^{m_2}, \dots, x_{r-1}^{n_{r-1}} = x_r^{m_r}, \, x_r^{n_r} = 1\}$$

is trivial.

Proof. By Lemma 2, $G_{m_1, n_1, \dots, m_r, n_r} = \{x_r : \ x_r^{m_1 \cdots m_r} = 1, \, x_r^{n_r} = 1\}$. Since $gcd(m_1 \cdots m_r, n_r) = 1$, the result follows. $\square$

Lemma 3. *Let H_r be the abelian group*

$$H_r = \{x_1, \dots, x_r : \ [x_i, x_j] = 1, \, x_1^{m_1} = 1, \, x_1^{n_1} = x_2^{m_2}, \dots, x_{r-1}^{n_{r-1}} = x_r^{m_r}, \, x_r^{n_r} = 1\}.$$

If $H_r = 1$ then $gcd(m_i, n_j) = 1$ for $1 \leq i \leq j \leq r$.

Proof. If $r = 1$, $H_1 = \{x_1 : \ x_1^{m_1} = 1, \, x_1^{n_1} = 1\} = 1$ implies that $gcd(m_1, n_1) = 1$. For $r > 1$, if $H_r = 1$, then $H_r / \langle x_r \rangle \cong H_{r-1} = 1$ and therefore by induction $gcd(m_i, n_j) = 1$ for $i \leq j \leq r-1$. It follows from Lemma 2, that the group $G_r = \{x_1, \dots, x_r : \ x_1^{m_1} = 1, \, x_1^{n_1} = x_2^{m_2}, \dots, x_{r-1}^{n_{r-1}} = x_r^{m_r}\} = \{x_r : x_r^{m_1 \dots m_r} = 1\}$. Then $H_r = G_r / \langle x_r^{n_r} \rangle = \{x_r : x_r^{m_1 \dots m_r} = 1, \, x_r^{n_r} = 1\} = 1$ implies that $gcd(m_1 \dots m_r, n_r) = 1$. $\square$

We know state the main theorem of this section.

Theorem 3. *For the* 2*-stratifold* $X = X(m_1, n_1, m_2, n_2, \dots, m_r)$ *the following are equivalent:*
(1) X *is simply connected*
(2) $H_1(X) = 0$
(3) $gcd(m_i, n_j) = 1$ *for* $1 \leq i \leq j \leq r$.

Proof. Clearly (1) implies (2). Now $H_1(X)$ is the group H_r of Lemma 3, hence (2) implies (3). Finally (3) implies (1) by Corollary 2. □

5. Trivalent 2-stratifolds

In this section a 2-stratifold X and its associated labeled bicolored graph G are defined to be *trivalent*, if each black vertex b is incident to either three edges each with label 1 or to two edges, one with label 1, the other with label 2, or b is a terminal vertex with adjacent edge of label 3. This means that a neighborhood of each component C of the 1-skeleton X_1 has 3 sheets, so the permutation $\pi : \{1,2,3\} \to \{1,2,3\}$ of the regular neighborhood $N(C) = N_\pi(C)$ has partition $1+1+1$ or $1+2$ or 3.

The figure below shows a simple example with one black vertex with partition $1+1+1$ and all terminal edges with label 2. A simple computation shows that the associated 2-stratifold has fundamental group $\mathbb{Z}_2$. In fact we show that in general the homology is non-trivial:

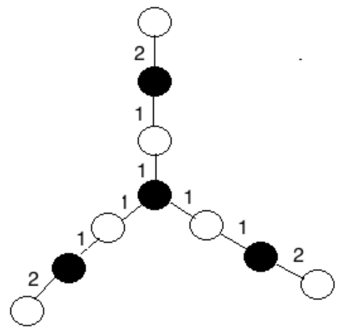


Figure 3: One branch vertex

Lemma 4. *If all terminal edges of a trivalent graph* G *(with a non-zero number of edges) have label* 2*, then* $H_1(X, \mathbb{Z}_2) \neq 0$.

Proof. By a *branch* vertex we mean a vertex of degree ≥ 3. A *terminal branch* is a connected linear subgraph L of G that joins a branch vertex v to a terminal vertex such that no vertex of $L - \{v\}$ is a branch vertex.

Note that all terminal vertices of G are white.

Suppose the Lemma is false. Let G be a counterexample with a smallest number $n > 0$ of edges. If w is a white branch vertex, then for the subcomplex $T = r^{-1}(\Gamma)$, where Γ is the complement of a component of $G - \{w\}$, there is a surjection $H_1(X, \mathbb{Z}_2) \to H_1(\hat{T}, \mathbb{Z}_2)$, which shows that Γ is a counterexample to the Lemma with fewer than n edges. Therefore G does not have white branch vertices.

We first claim that every terminal branch of G has length 3 with labels $2, 1, 1$ (starting from the (white) terminal vertex).

To see this, let $w_1 - b_1 - w_2 - b_2 - ... - w_r - b_r$ be a terminal branch where w_1 is the (white) terminal vertex and b_r is the (black) branch point. We must have $r \geq 2$ since the labels of $w_1 - b_1$ and $b_1 - w_2$ are 2 and 1 respectively so b_1 cannot be a branch vertex. Suppose the label of the edge $w_2 - b_2$ is 2. Then, eliminating the edges $w_1 - b_1 - w_2$ from G we obtain a counterexample with $n - 2$ edges. Hence the label of $w_2 - b_2$ is 1. If $r > 2$ then the label of $b_2 - w_3$ is 2 and for the linear subgraph $w_1 - b_1 - w_2 - b_2 - w_3$ of G the corresponding sub complex Y has $H_1(Y, \mathbb{Z}_2) = 0$, which a simple computation shows is not true. Therefore $r = 2$ and $b_r = b_2$ is a branch vertex with partition $1 + 1 + 1$, so the terminal branch $w_1 - b_1 - w_2 - b_2$ has length 3 and labels $2, 1, 1$.

Now, let b be an outermost branch vertex of G (that is a (black) terminal vertex of the tree obtained from G by removing all its terminal branches). If b is the only branch point of G then G is the graph in figure 3, for which the corresponding 2-stratifold has non-trivial $\mathbb{Z}_2$-homology, a contradiction. In any other case two terminal branches of G have b as an endpoint. Let Γ be the graph that is obtained from G by replacing the two terminal branches by a one-edge linear graph with end point b and label 2. The 2-stratifold corresponding to the union of the two branches has $\mathbb{Z}_2$ homology $= \mathbb{Z}_2$ and it follows that the 2-stratifold Y that corresponds to the graph Γ has the same $\mathbb{Z}_2$-homology group as X. Thus Γ is a counterexample with $n - 5$ edges, contradicting the minimality of n. □

In the proof of the following theorem we give an efficient algorithm to decide whether or not a trivalent 2-stratifold is 1-connected.

Theorem 4. *There is an algorithm for determining if a trivalent 2-stratifold X is 1-connected.*

Proof. By Corollary 1, G must be a tree, all white vertices have genus 0 and all terminal vertices are white.

Step (1) Delete all terminal branches of length 2 with labels $1, 2$ from G (starting from the (white) terminal vertex).

The fundamental group of X is not changed.

Step (2) If there is a black branch vertex b with a terminal (white) neighbor, delete b and its edges.

This splits G into two subgraphs Γ_1 and Γ_2 corresponding to two subcomplexes Y_1 and Y_2 of X such that $\pi_1(X) \cong \pi_1(Y_1) * \pi_1(Y_2)$. Then $\pi_1(X) = 1 \iff \pi_1(Y_1) = 1$ and $\pi_1(Y_2) = 1$.

Repeat steps (1) and (2) as long as possible for the resulting components Y_i. Then either some Y_i has all terminal edges of label 2 or the process yields a collection of white vertices. In the first case $\pi_1(Y_i) \neq 1$ by Lemma 4 and therefore $\pi_1(X) \neq 1$. In the second case $\pi_1(X) = 1$. $\square$

We finally give a homology classification of trivalent simply-connected 2-stratifolds.

Theorem 5. *A trivalent 2-stratifold is* 1*-connected if and only if* $H_1(X; \mathbb{Z}_6) = 0$.

Proof. Note that the condition is equivalent to $H_1(X; \mathbb{Z}_2) = 0$ and $H_1(X; \mathbb{Z}_3) = 0$. The first condition implies by Theorem 1(a),(b) that G is a tree with all white vertices of genus 0. Since G is trivalent, for a terminal black vertex T we would have a surjection $H_1(X; \mathbb{Z}_3) \to H_1(\hat{T}, \mathbb{Z}_3) \cong \mathbb{Z}_3$, which is impossible by the second condition. Thus all terminal vertices are white. Following the algorithm, we note that in step (1) the $\mathbb{Z}_2$-homology does not change and in step (2), $H_1(X; \mathbb{Z}_2) = H_1(Y_1; \mathbb{Z}_2) + H_1(Y_2; \mathbb{Z}_2)$. Thus we end up with a (possibly disconnected) graph with no edges. This implies that $\pi_1(X)$ is a free product of trivial groups, that is, $\pi_1(X) = 1$. $\square$

References

[1] G. Carlsson: *Topology and data.* Bull. AMS (New Series) **46 (2)** (2009) 255–308.

[2] G. Carlsson, T. Ishkhanov, V. de Silva, A. Zomorodian: *On the Local Behavior of Spaces of Natural Images.* International Journal of Computer Vision **76 (1)** (2008) 1–12.

[3] J. S. Carter J: *Reidemeister/Roseman-type moves to embedded foams in 4-dimensional space.* arXiv: **1210.3608v1** [math.GT]

[4] J. C. Gómez-Larrañaga, F. González-Acuña, Wolfgang Heil: *Categorical group invariants of 3-manifolds.* Manuscripta math. **145** (2014) 433–448.

[5] M. Kreck: *Differential Algebraic Topology: From Stratifolds to Exotic Spheres.* AMS (2010).

[6] M. Khovanov: *sl(3) link homology.* Algebr. and Geom. Topol. **4** (2004) 1045–1081.

[7] P. Lum, G. Singh, J. Carlsson, A. Lehman, T. Ishkhanov, M. Vejdemo-Johansson, M. Alagappan, G. Carlsson: *Extracting insights from the shape of complex data using topology*. Nature Scientific Reports **3** (2013) 12–36.

[8] E. Michael: *Collared Sets, in General Topology and its Relations to Modern Analysis and Algebra*. Proc. Symp. Prague (1962) 270–271.
http://dml.cz/dmlcz/700920

[9] J. Milnor: *Singular Points of Complex Hypersurfaces*. Annals of Mathematics Studies, Princeton University Press **61** (1969).

[10] W. Neumann: *A calculus for plumbing applied to the topology of complex surface singularities and degenerating complex curves*. Trans. AMS **268** (1981) 299–344.

[11] R. Thom: *Ensembles et morphismes stratifies*. Bull. AMS **75** (1969) 240–284.

[12] S. Weinberger: *The topological classification of stratified spaces*. Chicago Lectures in Mathematics. Chicago, IL, University of Chicago Press (1994).

Surface knot groups and 3-manifold groups

Francisco González-Acuña and Arturo Ramírez

CIMAT, Guanajuato México; Instituto de Matemáticas UNAM, Cuernavaca México
ficomx@yahoo.com.mx

CIMAT, Guanajuato México
ramirez@cimat.mx

Dedicado a José María Montesinos Amilibia, con cariño, en su septuagésimo aniversario.

ABSTRACT

We prove that the only orientable closed 3-manifold group which is the fundamental group of the complement of a smooth orientable closed surface in S^4 is $\mathbb{Z}$.

2010 Mathematics Subject Classification: 2000, Primary 57.
Key words: Surface knot groups, 3-manifold groups.

1. Introduction

The problem of determining which fundamental groups of spaces, in some interesting class of spaces, are 3-manifolds groups is often considered in the literature. For example in [DS] it is proved that the Kahler groups (that is, fundamental groups of compact Kahler manifolds) which are 3-manifolds groups are precisely the finite 3-manifolds groups which, by Perelman, are the finite subgroups of $\mathcal{O}(4)$ that act freely on S^3.

An interesting class of spaces is the collection of complements, in S^4, of orientable smooth, closed 2-submanifolds (see for example [6], [12], [15]). The class of fundamental groups of such complements is denoted by $\mathcal{S}$ in [6]. The groups of $\mathcal{S}$ are also called irreducible C-groups (see [12]). We prove in the present article that only the infinite cyclic group is a member of $\mathcal{S}$ that is the fundamental group of an orientable closed 3-manifold. A motivation for our interest in this question is a conjecture of J. Simon which we now state in a strengthned form. Let $(k;0)$ be a closed 3-manifold obtained by longitudinal surgery on a nontrivial knot k in S^3. Conjecture: $\pi_1(k;0)$

does not belong to $\mathcal{S}$. By Gabai [3] $\pi_1(k;0) \ncong \mathbb{Z}$ and so, it follows by our result that the conjecture is true.

A manifold will be called *closed* if it is compact, connected with empty boundary. A map $f : X \to Y$ between arcwise connected spaces is *essential* if $f_\# : \pi_1 X \to \pi_1 Y$ is injective. Arrows with no label are induced by inclusion.

The symbol $[a, b]$ is $aba^{-1}b^{-1}$, $<<>>$ will denote normal closure and $\mathcal{G}$ is the class of finitely presented groups.

2. Main result

Let $\mathcal{S}_n$ be the class of fundamental groups $\pi_1(S^{n+2} - F^n)$ where F^n is a smooth orientable closed n-submanifold of S^{n+2}.

It is known that $\mathcal{S}_1 \subsetneq \mathcal{S}_2$ and $\mathcal{S}_n = \mathcal{S}_2$ for $n \geq 2$, so we denote $\mathcal{S}_n$ by $\mathcal{S}$ if $n \geq 2$.

Let $\mathcal{M}_3$ be the class of fundamental groups of closed orientable 3-manifolds.

Theorem 2.1 $\mathcal{S} \cap \mathcal{M}_3 = \{\mathbb{Z}\}$.

Clearly $\pi_1(S^2 \times S^1) \simeq \mathbb{Z} \simeq \pi_1(S^4 - S^2)$, where S^2 is the trivial 2-knot in S^4, so $\mathbb{Z} \in \mathcal{S} \cap \mathcal{M}_3$.

Before giving a group-theoretical characterization of the class $\mathcal{S}$ we recall the following definition. Suppose a, $b \in G$ and $[a, b] = 1$; then the Pontrjagin product of a and b, which we denote by $a \wedge b$, is the image of the canonical generator of $H_2(\mathbb{Z} \times \mathbb{Z})$ under $H_2(\mathbb{Z} \times \mathbb{Z}) \stackrel{(\varphi_{a,b})_*}{\longrightarrow} H_2(G)$, where $\varphi_{a,b} : \mathbb{Z} \times \mathbb{Z} \to G$ is the homomorphism such that $\varphi_{a,b}(1, 0) = a$ and $\varphi_{a,b}(0, 1) = b$. If $t \in G$ and $\mathcal{C}_t$ is the centralizer of t in G, then we write $t \wedge \mathcal{C}_t = \{t \wedge c : c \in \mathcal{C}_t\}$.

Notice that if $\mathcal{C}_t$ is cyclic then $t \wedge \mathcal{C}_t = 0$ because $(\varphi_{a,b})_*$ factors through the trivial group $H_2(\mathcal{C}_t)$.

The following "intrinsic" (i.e. not involving presentations) characterization of the group $\mathcal{S}$ is a slight reformulation of a theorem of Simon [15], using a remark in [2].

Theorem 2.2 (Simon) $\mathcal{S} = \{G \in \mathcal{G} : H_1 G \simeq \mathbb{Z}$ *and there exists* $t \in G$ *such that* $\langle\langle t \rangle\rangle = G$ *and* $t \wedge \mathcal{C}_t = H_2(G)\}$.

One can also give the following characterization. A *Wirtinger presentation* is a finite presentation $\langle x_1, ..., x_m : r_1, ..., r_n\rangle$ such that each relator r_k is of the form $x_i^{-1} w^{-1} x_j w$. Groups having a Wirtinger presentation are also called LOG group (see)s Then (see [15])

Theorem 2.3 $\mathcal{S} = \{G \in \mathcal{G} : H_1 G \simeq \mathbb{Z}$ *and* G *has a Wirtinger presentation*$\}$.

A presentation $(x_1, ..., x_n; r_1, ..., r_n)$ such that $\prod\limits_{i=1}^{n} r_i x_i r_i^{-1} = \prod\limits_{i=1}^{n} x_i$ in the free group $F_n(x_1, ..., x_n)$ is called an *Artin* presentation.

From the fact that every closed orientable 3-manifolds is an open book with planar pages the following characterization of $\mathcal{M}_3$ can be given ([6], [5], [18]).

Theorem 2.4 *$\mathcal{M}_3$ is the class of groups in $\mathcal{G}$ that have an Artin presentation.*

Wall has proposed the following intrinsic characterization of $\mathcal{M}_3$.

Conjecture 2.5 *$G \in \mathcal{M}_3$ iff G is a free product of finitely many PD_3 groups.*

See [1, p. 19].

3. Proofs

Lemma 3.1 *Let $G = \pi_1 M^3$ where M^3 is closed, $H_1 M^3 = \mathbb{Z}$ and G is normally generated by one element. Then M^3 is prime.*

Proof. If M^3 is not prime, then $M^3 = M_1^3 \# \Sigma_1^3 \# ... \# \Sigma_n^3$ $(n \geq 1)$, $H_1 M_1^3 = \mathbb{Z}$, M_1 is prime and Σ_1^3,...,Σ_n^3 are prime homology spheres whith $\pi_1 \Sigma_i^3 \neq 1$ $(i = 1, ..., n)$ (by Perelman). Note that $\pi_1 M_1^3 * \pi_1 \Sigma_1^3$, and therefore $\mathbb{Z} * \pi_1 \Sigma_1^3$, is a quotient of G and so, it can be normally generated by one element. Since Σ_1^3 is prime $\pi_1 \Sigma_1^3$ is finite or torsionfree [8]. In both cases $\mathbb{Z} * \pi_1 \Sigma_1^3$ can not be normally generated by one element [7], [11]. Therefore M^3 is prime. □

Lemma 3.2 *Let W^3 be a compact orientable 3-manifold, such that $\partial W^3 \neq \varnothing$ and $\dim H_1(\partial W^3; \mathbb{Q}) \geq 2 \cdot \dim H_1(W^3; \mathbb{Q})$. Then $H_2(\partial W^3; \mathbb{Q}) \twoheadrightarrow H_2(W^3; \mathbb{Q})$ is surjective.*

Proof. By Poincare duality $\dim H_2(W^3, \partial W^3; \mathbb{Q}) = \dim H_1(W^3; \mathbb{Q})$, and so

$$H_1(\partial W^3; \mathbb{Q}) \geq \dim H_1(W^3; \mathbb{Q}) + \dim H_2(W^3, \partial W^3; \mathbb{Q})$$

and therefore in the exact sequence

$$H_2(\partial W^3; \mathbb{Q}) \to H_2(W^3; \mathbb{Q}) \to H_2(W^3, \partial W^3; \mathbb{Q}) \to H_1(\partial W^3; \mathbb{Q}) \to H_1(W^3; \mathbb{Q})$$

the third arrow is injective and the first arrow is surjective. □

Corollary 3.1 *Let W^3 be a compact orientable Seifert 3-manifold, with $\partial W^3 \neq \varnothing$, whose orbit space is orientable of genus 0 or nonorientable of genus 1. Then the image of $H_2(\partial W^3) \to H_2(W^3)$ has finite index.*

Proof. If m is the number of boundary components of ∂W^3 it is easy to see that $H_1(W^3; \mathbb{Q}) = m$. □

Lemma 3.3 *Let $M^3 \to F^2$ be a Seifert fibration with n $(n \geq 0)$ exceptional fibers. Suppose that M^3 is orientable, closed, $H_1 M^3 \simeq \mathbb{Z}$ and $M^3 \not\cong S^2 \times S^1$. Then $F^2 \simeq S^2$ and $n > 0$.*

Proof. From the usual presentation of $\pi_1 M^3$ (see [14, §1]), one sees that $\pi_1 F^2$ is a quotient of $\pi_1 M^3$. Hence $H_1 F^2$ is a quotient of $H_1 M^3$ and so $H_1 F^2$ is cyclic, which implies that F^2 is S^2 or P^2. If $F^2 \simeq P^2$

$$\pi_1 M^3 = \big|\big|q_1, ..., q_n, h, v : q_i^{\alpha_i} h^{\beta_i} = vhv^{-1}h = q_1...q_n v^2 h^{-b} = 1\big|\big|$$

From this one sees that $h, q_1, ..., q_n$ and v have finite order in $H_1 M^3$ and so $H_1 M^3$ is finite, contradicting $H_1 M^3 \simeq \mathbb{Z}$. Therefore $F^2 \simeq S^2$ and

$$\pi_1 M^3 \simeq \big|\big|q_1, ..., q_n, h : [h, q_i] = q_i^{\alpha_i} h^{\beta_i} = q_1...q_n h^{-b} = 1\big|\big|$$

If $n = 0$, then as $H_1 M^3 \simeq \mathbb{Z}$, we must have $b = 0$ and $M^3 \simeq S^2 \times S^1$; hence $n > 0$. □

Lemma 3.4 *Let M^3 be an orientable closed irreducible 3-manifold with $H_1(M^3) = \mathbb{Z}$. Assume that M^3 is not homeomorphic to the torus bundle over S^1 with monodromy of order 6. Then any essential map from the torus T^2 into M^3 is homotopic to a map whose image is contained in a Seifert space with non empty boundary.*

Proof. By [10], [9] or [16] the assertion is true if M^3 is not a Seifert space.

We can therefore assume that M^3 is a Seifert space. By Lemma 3.3, M^3 has S^2 as orbit space and n $(n > 0)$ exceptional fibers. Let e_1,..., e_n be the exceptional fibers and

$$W^3 = \overline{M^3 - \cup_{i=1}^n N(e_i)}$$

where the solid tori $N(e_1)$,..., $N(e_n)$ are disjoint saturated tubular neighborhoods of e_1,..., e_n. We will prove that any map from the torus T^2 into M^3 is homotopic to a map whose image is contained in W^3.

We can identify W^3 with $S_0^2 \times S^1$ where S_0^2 is the 2-sphere with n holes. and $\pi_1 S_0^2 = ||q_1, ..., q_n : q_1...q_n = 1||$ and so

$$\pi_1 W^3 = ||q_1, ..., q_n, h : q_1...q_n = [q_i, h] = 1||$$

where h corresponts to a generator of S^1. We can write

$$\pi_1 M^3 = \big|\big|q_1, ..., q_n, h : [q_i, h] = q_i^{\alpha_i} h^{\beta_i} = q_1...q_n h^{-b} = 1\big|\big|$$

and the inclusion-induced epimorfism $\varphi : \pi_1 W^3 \longrightarrow \pi_1 M^3$ sends q_i to q_i and h to h.

Consider $\pi_1 M^3 / < h >$, the quotient of $\pi_1 M^3$ by the central subgroup generated by h. We identify it with $||q_1, ..., q_n : q_1...q_n = q_i^{\alpha_i} = 1||$; the natural epimorphism $\rho : \pi_1 M^3 \to \pi_1 M^3 / < h >$ sends q_i to q_i and h to 1. We have the epimorphism $\psi : \pi_1 S_0^2 \to \pi_1 M^3 / < h >$ where $\psi(q_i) = q_i$ and the epimorphism $p : \pi_1 W^3 \to \pi_1 S_0^2$ where $p(q_i) = q_i$ and $p(h) = 1$. Define also the homomorphism $\sigma : \pi_1 S_0^2 \to \pi_1 W^3$ by $\sigma(q_i) = q_i$; thus $p \circ \sigma$ is the identity. We have the commutative diagram

$$\begin{array}{ccc} & & \ker p = < h > \\ & & \downarrow \\ \pi_1 W^3 & \xrightarrow{\varphi} & \pi_1 M^3 \\ \sigma \uparrow\downarrow p & \circlearrowleft & \downarrow \rho \\ \pi_1 S_0^2 & \xrightarrow{\psi} & \pi_1 M^3 / < h > \end{array}$$

Claim: If γ_1, $\gamma_2 \in \pi_1 M^3$, and $[\gamma_1, \gamma_2] = 1$, then there exist u_1, $u_2 \in \pi_1 W^3$ such that $[u_1, u_2] = 1$ and $\varphi(u_i) = \gamma_i$ $(i = 1, 2)$.

Proof of the claim. By [13, Proposition 7.10], if we exclude the 7 cases in which the orbifold Euler characteristic $\chi - \Sigma(1 - \frac{1}{\alpha_i})$ is zero, the group, $\pi_1 M^3 / < h >$ is Fuchsian and every Abelian subgroup of it is cyclic. Six of these seven cases do not arise because $||q_1, ..., q_n : q_1...q_n = q_i^{\alpha_i} = 1||$ has non-cyclic Abelianization which is not possible since $H_1 M^3 \simeq \mathbb{Z}$. The remaining case ($\alpha_1 = 2$, $\alpha_2 = 3$,$\alpha_3 = 6$) also does not arise because, by hypotesis, M is not the Seifert space $(Oo0; -1, (2,1), (3,1), (6,1))$, the torus bundle over S^1 with monodromy of order 6. Hence every Abelian subgroup of $\pi_1 M^3 / < h >$ is cyclic.

Let γ_1 and γ_2 be commuting elements of $\pi_1 M^3$. Since $< \rho(\gamma_1), \rho(\gamma_2) >$ is cyclic, there exist $w \in \pi_1 S_o^2$ and integers r_1 and r_2 such that $\psi(w)^{r_i} = \rho(\gamma_i)$ $(i = 1, 2)$.

Let $w_i' = \sigma(w)^{r_i}$. Clearly $[w_1', w_2'] = 1$, $\rho \circ \varphi(w_i') = \psi(w_i') = \rho(\gamma_i)$, and so $\gamma_i = \varphi(w_i')h^{e_i} = \varphi(w_i' h^{e_i})$, for some integer e_i. Writing $u_i = w_i' h^{e_i}$, we have $[u_1, u_2] = 1$ and $\varphi(u_i) = \gamma_i$ $(i = 1, 2)$. This proves the claim.

Now let $f : T^2 \to M^3$ be a map and let γ_1, $\gamma_2 \in \pi_1 M^3$ be images under $f_\#$ of two generators τ_1 and τ_2 of $\pi_1 T^2$. Clearly $[\gamma_1, \gamma_2] = 1$. Then by the claim there exist u_1, $u_2 \in \pi_1 W^3$ such that $[u_1, u_2] = 1$ and $\iota_\#(u_i) = \gamma_i$ $(i = 1, 2)$, where $\iota : W^3 \to M^3$ is the inclusion. Since W^3 is aspherical there is a map $g : T^2 \to W^3$ such that $g_\#(\tau_i) = u_i$ $(i = 1, 2)$. Then $(i \circ g)_\# = f_\#$ and, since M^3 is aspherical, $i \circ g$ is homotopic to f. □

Proposition 3.1 *Let M^3 be a closed orientable irreducible 3-manifold with $H_1(M^3) = \mathbb{Z}$ and let $f : T^2 \to M^3$ be continuous. Assume that $M^3$is not homeomorphic to the torus bundle over S^1 with monodromy of order* 6. *If $f_* : H_1(T^2) \to H_1(M^3)$ is nonzero then $f_* : H_2(T^2) \to H_2(M^3)$ is zero.*

Proof. If f is not essential then there is a nontrivial loop α in T^2 such that $f \circ \alpha$ is trivial in $\pi_1 M^3$. Since $\pi_1 M^3$ is torsion free, because M^3 is aspherical, we can take α primitive. Then $f_\# : \pi_1 T^2 \to \pi_1 M^3$ factors through $\pi_1 M^3 / < h >= \mathbb{Z}$ and we have a commutative triangle

$$\begin{array}{ccccc} H_2(\pi_1 T^2) & & \xrightarrow{(f_\#)_*} & & H_2(\pi_1 M^3) \\ & \searrow & \circlearrowleft & \nearrow & \\ & & H_1 \mathbb{Z} = 0 & & \end{array}$$

Hence $H_2 T^2 \xrightarrow{f_*} H_2 M^3$ is zero.

We may therefore assume f is essential.

By Lemma 3.4 we may assume there is a 3-submanifold W^3 of M^3, with nonempty boundary, such that $f(T^2) \subset W^3$ and W^3 is a Seifert manifold.

Since $f_* : H_1(T^2) \to H_1(W^3)$ is nonzero, $H_1(W^3) \to H_1(M^3)$ is nonzero. Let $E^3 = \overline{M^3 - W^3}$ .From the exact sequence

$$H_1W^3 \to H_1M^3 \to H_1(M^3, W^3) \to \widetilde{H}_0W^3$$

one sees that $H_1(M^3, W^3)$ is finite cyclic and, since $H_1(E^3, \partial E^3) \simeq H_1(M^3, W^3)$ by excision, the middle arrow in

$$H_1(E^3, \partial E^3) \to H_0(\partial E^3) \to H_0E^3 \to H_0(E^3, \partial E^3)$$

is an isomorphism. Hence every component of E^3 has connected boundary.

Let γ be the union of $n-1$ disjoint arcs properly embeded in W^3 such that $\gamma \cup \partial W^3$ is connected. We have, using excision and the homology sequence of $(W^3, \partial W^3)$ and $(M^3, \gamma \cup E^3)$

$$\begin{aligned}\operatorname{coker}(H_1(\partial W^3) \to H_1W^3) &\simeq \operatorname{coker}(H_1(\gamma \cup \partial W^3) \to H_1W^3)\\ &\simeq H_1(W^3, \gamma \cup \partial W^3) \simeq H_1(M^3, \gamma \cup E^3)\\ &\simeq \operatorname{coker}(H_1(\gamma \cup E^3) \to H_1M^3)\end{aligned}$$

which is cyclic.

If the orbit space of W^3 is orientable of genus g, $\operatorname{coker}(H_1(\partial W^3) \to H_1W^3) \simeq \mathbb{Z}^{2g}$ so $g = 0$, and if it is nonorientable of genus k, $\operatorname{coker}(H_1(\partial W^3) \to H_1W^3) \simeq \mathbb{Z}_2 \oplus \mathbb{Z}^{k-1}$ so $k = 1$. Therefore by Corollary 3.1 the image of $H_2(\partial W^3) \to H_2W^3$ has finite index.

Write $f = j \circ g$ where $g : T^2 \to W^3$ and $j : W^3 \to M^3$ is the inclusion and also denote the inclusion of ∂W^3 in W^3 by i. Notice that $j_*i_* : H_2(\partial W^3) \to H_2(M^3)$ is zero because every component of ∂W^3 bounds in M^3.

Now if $[T^2]$ is a generator of $H_2(T^2)$, $r \cdot g_*([T^2]) \in \operatorname{im} i_*$ for some positive integer r, because $\operatorname{im} i_*$ has finite index. Hence $j_*(r \cdot g_*([T^2])) = rf_*([T^2])) = 0$ and therefore $f_*([T^2])) = 0$ since $H_2(M^3) \simeq \mathbb{Z}$. □

Proposition 3.2 *Let M^3 be the torus bundle over S^1 whith monodromy $\varphi : S^1 \times S^1 \to S^1 \times S^1$ of order 6, that is $\varphi(z_1, z_2) = (z_2, z_1^{-1}z_2)$, where $S^1 = \{z \in \mathbb{Z} : |z| = 1\}$. Let $m \in \pi_1(M^3)$ be such that $<< m >>= \pi_1(M^3)$. Then $\mathcal{C}_m$, the centralizer of m in π_1M^3, is $< m >$, and so $m \wedge \mathcal{C}_m = \{0\}$.*

Proof. We have $\pi_1M^3 = \left\|a, b, t : [a, b], t^{-1}at = b, t^{-1}bt = a^{-1}b\right\|$ where the center is generated by t^6 and the commutator subgroup is $< a, b >$, the free Abelian group of rank two. The monodromy induces an isomorphism $\varphi :< a, b > \to < a, b >$ where $\varphi(a) = b$ and $\varphi(b) = a^{-1}b$ with corresponing matrix $\begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$. Notice that the only fixed point of φ is the trivial element. □

Consider now an element m in $\pi_1 M^3$ whose normal closure is $\pi_1 M^3$; we can assume this $m = tc$ with $c \in < a, b >$. We claim that $\mathcal{C}_{\pi_1 M^3}(tc) =< m >$. Indeed, if $[t^e c_1, tc] = 1$ with $c_1 \in < a, b >$ then $t^e c_1 = c_2(tc_1)^e$ for some $c_2 \in < a, b >$ and then $1 = [c_2(tc_1)^e, tc] = c_2 t c_2^{-1} t^{-1} = c_2 \varphi^{-1}(c_2^{-1})$, hence $\varphi(c_2) = c_2$ so $c_2 = 1$. Therefore $t^e c_1 = (tc)^e$ and $\mathcal{C}_{\pi_1 M^3}(m) =< m >$. $\square$

Proof of Theorem 2.1. Let $G = \pi_1(M^3)$, with M^3 closed. Suppose $G \simeq \pi_1(S^4 - F^2)$ where F^2 is a closed, smooth 2-submanifold of S^4. Suppose $G \neq \mathbb{Z}$. Let $m \in G$ be such that $\langle\langle m \rangle\rangle = G$ and $m \wedge \mathcal{C}_m = H_2(G)$ where $\mathcal{C}_m = \{g : [m, g] = 1\}$; m represents a generator of $H_1 G (\simeq H_1 M^3 \simeq \mathbb{Z})$. By Lemma 3.1 M^3 is irreducible and therefore is aspherical [8], and so $H_i G \simeq H_i M^3$ $(\forall i)$.

Assertion: $m \wedge \mathcal{C}_m = \{0\}$.

Proof of the assertion. By Propostion 3.2, we may assume that M^3 is not homeomorphic to the torus bundle with monodromy of order 6.

Let $c \in \mathcal{C}_m$. Let φ be the homomorphism from $\mathbb{Z}^2 (= \pi_1(S^1 \times S^1))$ to $G (= \pi_1 M^3)$ such that $\varphi(1, 0) = c$ and $\varphi(0, 1) = m$. Since M^3 is aspherical there is a map $f : S^1 \times S^1 \to M^3$ such that $f_* = \varphi$. As $[\varphi(1, 0)]$ generates $H_1 M^3$, by Proposition 3.1, $f_*([S^1 \times S^1]) = 0 \in H_2 M^2$, that is, $m \wedge c = 0$.This proves the assertion.

Since $H_2 G \simeq H_2 M^3 \simeq H_1 M^3 \simeq \mathbb{Z}$, by Poincare duality, $m \wedge \mathcal{C}_m \neq H_2 G$. Therefore $G \simeq \mathbb{Z}$. $\square$

Theorem 2.1 is not valid if we redefine $\mathcal{M}_3$ so as to include fundamental groups of the closed nonorientable 3-manifolds: the manifolds obtained by blowing up a non-trivial knot in S^3 have a non-cyclic fundamental group that belongs to $\mathcal{S}$.

References

[1] L. Bessiéres, G. Besson, S. Maillot, M. Boileau, J. Porti: *Geometrisation of 3-manifolds.* EMS Tracts in Mathematics 13, European Mathematical Society (EMS), Zürich 2010.

[2] F. R. Beyl, J. Tappe, *Group extensions, representations, and the Schur multiplicator.* LNM 958, Springer-Verlag, Berlin-New York 1982.

[3] D. Gabai, *Foliations and the topology of 3-manifolds II.* J. Differential Geom. 26 (1987) 461–478.

[4] N. D. Gilbert, J. Howie: *LOG Groups and cyclically presented groups.* J. Algebra 174 (1995) 118–131.

[5] F. González-Acuña: *Open books.* Lecture Notes, University of Iowa, 1975, 12 pp. http://www.oberlin.edu/faculty/jcalcut/ga.pdf

[6] F. González-Acuña, C. McA. Gordon, J. Simon: *Unsolvable problems about higher-dimensional knots and related groups.* L'Enseignement Mathématique 56, Issue 1/2 (2010) 143–171.

[7] M. Gerstenhaber, O. S. Rothaus: *The solution of sets of equations in groups.* Proceedings of the National Academy of Sciences. 10, 48(9)(1962) 1531-3.

[8] J. Hempel: *3-manifolds.* Princeton University Press, Annals of Mathematical Studies, 86, 1976.

[9] W. Jaco, P. B. Shalen: *Seifert fibred spaces in 3-manifolds.* Memories AMS 220 (1979).

[10] K. Johanson: *Homotopy equivalences of 3-manifolds with boundary.* LNM 761,Springer-Verlag, Berlin 1979.

[11] A. Klyachko: *Funny property of sphere and equations over groups.* Comm. in Alg. 21(7) (1993) 2555–2575.

[12] V. S. Kulikov: *Alexander polynomials of plane algebraic curves.* Izv. Math. 42:1 (1994) 67–90.

[13] R. Lyndon, P. Schupp: *Combinatorial Group Theory.* Springer-Verlag, New York-Berlin 1977.

[14] P. Orlik, E. Vogt, H. Zieschang: *Zur topologie gefaserter dreidimensionaler mannigfaltigkeiten..* Topology 6 (1967) 49–64

[15] J. Simon: *Wirtinger approximations and the knot groups of F_n in S^{n+2}.* Pacific J. Math. 90 (1980) 177–190.

[16] P. Scott: *Strong annulus and torus theorems and the enclosing property of characteristic submanifolds of 3-manifolds.* Quart. J. Math. Oxford Ser. (2) 35 (1984), no. 140, 485–506.

[17] A. Dimca, A. I. Suciu: *Which 3-manifold groups are Kahler groups?* arXiv:0709.4350.

[18] H. E. Winkelnkemper: *Artin Presentations I: Gauge Theory, 3+1 TQFTs and the Braid Groups.* J. Knot Theory Ramifications, 11 (2) (2002) 223–275.

Some geometric properties of variable exponent Lebesgue spaces

Francisco L. Hernández* and **César Ruiz***

Departamento de Análisis Matemático
Facultad de Matemáticas
Universidad Complutense de Madrid
28040 Madrid, Spain
pacoh@mat.ucm.es Cesar_Ruiz@mat.ucm.es

Dedicado al profesor José María Montesinos Amilibia.

ABSTRACT

In this paper we survey some recent results on the geometry and the isomorphic structure of variable exponent Lebesgue spaces $L^{p(\cdot)}(\Omega, \Sigma, \mu)$. Averaging and orthonormal projections on these spaces are also studied.

2010 Mathematics Subject Classification: 46E30.
Key words: Variable exponent spaces, l_q-isomorphic copies, averaging projections.

1. Introduction.

The goal of this paper is to survey some results on the geometry and the structure of variable exponent Lebesgue spaces $L^{p(\cdot)}(\Omega, \Sigma, \mu)$. The last two decades have seen a strong interest in these spaces $L^{p(\cdot)}(\Omega)$ motivated for their successful use in some areas as Harmonic Analysis and P.D.E. (cf. [17],[3], [1]). Variable exponent Lebesgue spaces (or Nakano spaces) $L^{p(\cdot)}(\Omega)$ belong to the general class of modular spaces and Musielak-Orlicz spaces (cf. [14]). An important difference with classical Lebesgue L^p-spaces is that $L^{p(\cdot)}(\Omega)$ spaces are not rearrangement invariant.

*Partially supported by grant MTM2012-31286

2. Notation and Preliminaries.

Throughout the paper (Ω, Σ, μ) is a σ-finite separable non-atomic measurable space and $L_0(\Omega)$ is the space of all real measurable function classes. Given a μ-measurable function $p : \Omega \to [1, \infty)$, the *Variable Exponent Lebesgue space* (or Nakano space) $L^{p(\cdot)}(\Omega)$, is defined by the set of all measurable scalar function classes $f \in L_0(\Omega)$ such that the modular $\rho_{p(\cdot)}(f/r)$ is finite for some $r > 0$, where

$$\rho_{p(\cdot)}(f) = \int_\Omega |f(t)|^{p(t)}\, d\mu(t) < \infty.$$

And the associated Luxemburg norm is

$$||f||_{p(\cdot)} := \inf\{r > 0 : \ \rho_{p(\cdot)}(f/r) \leq 1\}.$$

With the usual pointwise order, $\left(L^{p(\cdot)}(\Omega), || \,||_{p(\cdot)}\right)$ is a Banach lattice.

Properties of these kind of spaces dependent of properties of the measurable exponent functions $p(\cdot)$. We write

$$p^- := \text{ess }\inf\{p(t) : t \in \Omega\} \qquad \text{and} \qquad p^+ := \text{ess }\sup\{p(t) : t \in \Omega\}.$$

By $p^+_{|B}$ and $p^-_{|B}$ we denote the essential supremum and infimum of the function $p(\cdot)$ over a measurable subset B. The conjugate function $p^*(\cdot)$ of $p(\cdot)$ is defined by the equation $\dfrac{1}{p(t)} + \dfrac{1}{p^*(t)} = 1$ almost everywhere $t \in \Omega$.

The topological dual of the space $L^{p(\cdot)}(\Omega)$, with $p^+ < \infty$, is the variable exponent space $L^{p^*(\cdot)}(\Omega)$. When $\mu(\Omega) < \infty$ and $p_1(t) \leq p_2(t)$ a.e., we have $L^{p_2(\cdot)}(\Omega) \subseteq L^{p_1(\cdot)}(\Omega)$ and this inclusion is bounded.

An space $L^{p(\cdot)}(\Omega)$ is separable if and only if $p^+ < \infty$. This is also equivalent to $L^{p(\cdot)}(\Omega)$ does not have any subspace isomorphic to l_∞.

An space $L^{p(\cdot)}(\Omega)$ is reflexive if and only if $1 < p^- \leq p^+ < \infty$. This is also equivalent to $L^{p(\cdot)}(\Omega)$ be uniformly convex. The spaces $L^{p(\cdot)}(\Omega)$ are p^--convex and p^+-concave Banach lattices (c.f. [10] , [3] [1]). See also [13] for other geometric properties like the Radon-Nikodym property and the Daugavet property.

The *essential range* of the exponent function $p(\cdot)$ is defined by

$$R_{p(\cdot)} := \{q \in [1, \infty) : \quad \forall \epsilon > 0 \quad \mu(p^{-1}(q - \epsilon, q + \epsilon)) > 0\}.$$

Note that the essential range is a closed subset of $[1, \infty)$; in particular, $R_{p(\cdot)}$ is compact when $p(\cdot)$ is essentially bounded. Clearly the values p^- and p^+ are always in the set $R_{p(\cdot)}$.

3. Results.

From now on, we consider separable $L^{p(\cdot)}(\Omega)$ spaces to give results on the existence of ℓ_q-subspaces for $1 \leq q < \infty$.

Theorem 3.1 *([6]) Let $L^{p(\cdot)}(\Omega)$ be separable. Then $L^{p(\cdot)}(\Omega)$ has a lattice-isomorphic copy of l_q if and only if $q \in R_{p(\cdot)}$.*

A direct consequence is the following result.

Proposition 3.1 *Let $L^{p(\cdot)}(\Omega)$ and $L^{p_1(\cdot)}(\Omega')$ be separable variable exponent spaces. If $L^{p(\cdot)}(\Omega)$ is lattice-isomorphic to $L^{p_1(\cdot)}(\Omega')$, then $R_{p(\cdot)} = R_{p_1(\cdot)}$.*

Also it follows the determination of the Rademacher type and cotype (see also [9]).

Proposition 3.2 *An $L^{p(\cdot)}(\Omega)$ space has type q (for $1 \leq q \leq 2$) if and only if $q \leq p^-$. And $L^{p(\cdot)}(\Omega)$ has cotype q (for $2 \leq q \leq \infty$) if and only if $p^+ \leq q$.*

We pass now to consider arbitrary isomorphisms. First let us remark that if $L^{p(\cdot)}(\Omega)$ is isomorphic to $L^1(\Omega)$, then $L^{p(\cdot)}(\Omega) = L^1(\Omega)$ (up to renorming). Indeed, assume that $L^{p(\cdot)}(\Omega) \neq L^1(\Omega)$. Then $1 < p^+ < \infty$. Consider $r > 0$ and a subset Ω' of $p^{-1}([p^- + r, p^+])$ with finite measure. Then

$$L^{p^+}(\Omega') \subseteq L^{p_{|\Omega'}(\cdot)}(\Omega') \subseteq L^{p^- + r}(\Omega'),$$

where $L^{p_{|\Omega'}(\cdot)}(\Omega')$ is a complemented subspace of $L^{p(\cdot)}(\Omega)$ (clearly $P(f) = f\chi_{\Omega'}$ is a continuous projection). Now, considering the sequence of the Rademacher functions in $L^{p_{|\Omega'}(\cdot)}(\Omega')$ we get that $L^{p_{|\Omega'}(\cdot)}(\Omega')$ has a closed complemented subspace isomorphic to l_2. Hence the space $L^{p(\cdot)}(\Omega) = L^1(\Omega)$ has also a complemented l_2-copy, which is a contradiction (cf. [12] Theorem 2.b.4).

Using the above, let us show that Corollary 3.1 does not hold for isomorphisms:

Let $L^{p(\cdot)}[0,2]$ be with $p(t) = 2$, if $t \in [0,1]$ and $p(t) = 2 + t = p_1(t)$, if $t \in (1,2]$. Then $L^{p_1(\cdot)}(1,2] \simeq l_2 \oplus H$, for some closed subspace H, and

$$L^{p(\cdot)}[0,2] = L^{p(\cdot)}[0,1] \oplus L^{p(\cdot)}(1,2] \simeq l_2 \oplus L^{p_1(\cdot)}(1,2]$$

$$\simeq l_2 \oplus l_2 \oplus H \simeq l_2 \oplus H \simeq L^{p_1(\cdot)}(1,2]\,.$$

Thus the spaces $L^{p(\cdot)}[0,2]$ and $L^{p_1(\cdot)}(1,2]$ are isomorphic, but $R_{p(\cdot)} = \{2\} \cup [3,4] \neq R_{p_1(\cdot)}$.

We continue studying isomorphic l_q-copies in $L^{p(\cdot)}(\Omega)$ spaces for $q < 2$. Let us consider a $L^{p(\cdot)}(\Omega)$ space, with $1 \leq p^- < 2$. Then $L^{p(\cdot)}(\Omega)$ has an isomorphic copy of l_q for every $q \in (p^-, 2]$. Indeed, take a natural number n, with $p^- + \frac{1}{n} < q$, and a measurable subset Ω_n , with $0 < \mu(\Omega_n) < \infty$, such that

$$L^{p^- + \frac{1}{n}}(\Omega_n) \subseteq L^{p_{|\Omega_n}(\cdot)}(\Omega_n) \subseteq L^1(\Omega_n).$$

Now, consider a sequence of q-stable independent random variables in $L^{p^- + \frac{1}{n}}(\Omega_n)$ (cf.[12]). Then, using Proposition IV.4.10 in [5], we deduce that $L^{p_{|\Omega_n}(\cdot)}(\Omega_n)$ has an isomorphic copy of l_q.

Theorem 3.2 *([6]) Let $L^{p(\cdot)}(\Omega)$ be with $p^+ < \infty$.*

(a) If $1 \le p^- \le 2$, then l_q is isomorphic to a closed subspace of $L^{p(\cdot)}(\Omega)$ if and only if $q \in R_{p(\cdot)} \cup [p^-, 2]$.

(b) If $p^- > 2$, then l_q is isomorphic to a closed subspace of $L^{p(\cdot)}(\Omega)$ if and only if $q \in R_{p(\cdot)} \cup \{2\}$.

Using that the essential range of the conjugate function p^* verifies that

$$R_{p^*(\cdot)} = \{r \ge 1 : \frac{1}{r} + \frac{1}{q} = 1 \text{ for } q \in R_{p(\cdot)} \},$$

we deduce the following.

Proposition 3.3 *Let $L^{p(\cdot)}(\Omega)$ and $L^{p_1(\cdot)}(\Omega')$ be reflexive. If $L^{p(\cdot)}(\Omega)$ is isomorphic to $L^{p_1(\cdot)}(\Omega')$ then $R_{p(\cdot)} \setminus \{2\} = R_{p_1(\cdot)} \setminus \{2\}$.*

Indeed, if $2 < q \in R_{p(\cdot)}$ then l_q is isomorphic to a subspace of $L^{p(\cdot)}(\Omega) \simeq L^{p_1(\cdot)}(\Omega')$, so we have $q \in R_{p_1(\cdot)}$. Assume now $1 < q < 2$ and $q \in R_{p(\cdot)}$, then $2 < q' \in R_{p^*(\cdot)}$ for $1/q' + 1/q = 1$. Hence $q' \in R_{p_1^*(\cdot)}$ and $q \in R_{p_1(\cdot)}$.

We pass now to study the existence of *complemented* subspaces which are isomorphic to l_q in spaces $L^{p(\cdot)}(\Omega)$. A duality argument and results above allow to consider only scalars q in the essential range set $R_{p(\cdot)}$. From Theorem 3.1, we know that a lattice-isomorphic copy of l_q in $L^{p(\cdot)}(\Omega)$, for $q \in R_{p(\cdot)}$, can be obtained considering the span of a suitable sequence of disjoint normalized functions of the form $\left(\frac{\chi_{A_n}(t)}{\mu(A_n)^{\frac{1}{p(t)}}}\right)$. Hence, we study the boundedness of the associated *orthogonal* projections T_A, defined by

$$T_A(f)(t) = \sum_{n=1}^{\infty} \left(\int_{A_n} \frac{f(s)}{\mu(A_n)^{\frac{1}{p^*(s)}}} d\mu(s) \right) \frac{\chi_{A_n}(t)}{\mu(A_n)^{\frac{1}{p(t)}}},$$

where $\frac{1}{p(t)} + \frac{1}{p^*(t)} = 1$ almost everywhere. Clearly, the sequences $\left(\frac{\chi_{A_n}(t)}{\mu(A_n)^{\frac{1}{p(t)}}}\right)$ and $\left(\frac{\chi_{A_n}(t)}{\mu(A_n)^{\frac{1}{p^*(t)}}}\right)$ are bi-orthogonal.

First consider the following illustrative case. Assume that the exponent function $p(\cdot)$ is of the form $p(t)\chi_{A_n}(t) = q_n$, for some scalar sequence $(q_n) \subset [1,\infty)$ and a disjoint measurable subset sequence (χ_{A_n}). Then, the orthogonal projection T_A associated to the sequence $\left(\dfrac{\chi_{A_n}(t)}{\mu(A_n)^{\frac{1}{p(t)}}}\right)$ is a bounded operator in $L^{p(\cdot)}(\Omega)$ which coincides with the *averaging operator* P_A defined by

$$P_A(f) = \sum_{n=1}^{\infty} \frac{\int_{A_n} f(s)d\mu(s)}{\mu(A_n)} \ \chi_{A_n}.$$

Indeed, by Jensen integral inequality, we get $\rho_{p(\cdot)}(T_A f) = \rho_{p(\cdot)}(P_A f) \leq \rho_{p(\cdot)}(f)$, for all $f \in L^{p(\cdot)}(\Omega)$.

In the general case the following estimation holds, if $(A_n)_{n=1}^{\infty}$ is a sequence of disjoint measurable sets with $\sum\limits_{n=1}^{\infty} \mu(A_n) \leq 1$, then the associated orthogonal projection T in $L^{p(\cdot)}(\bigcup_{n=1}^{\infty} A_n)$ verifies

$$\rho_{p(\cdot)}(\frac{Tf}{K}) \ \leq \ \sum_{n=1}^{\infty} \left(\int_{A_n} \left(\frac{|f|}{K}\right)^{p^-_{|A_n}} d\mu\right) \left(\frac{1}{\mu(A_n)}\right)^{p^+_{|A_n} - p^-_{|A_n}},$$

for some constant $K \geq 1$ and every normalized function f.

It turns out that given $p : \Omega \to [1,\infty)$ for each value $q \in R_{p(\cdot)}$ there exist a sequence of disjoint measurable sets (A_n) with $\sum_{n=1}^{\infty} \mu(A_n) \leq 1$, and a constant $M_q > 0$ verifying that the sequences $(p^-_{|A_n})_n$ *and* $(p^+_{|A_n})_n$ converge to q and

$$\left(\frac{1}{\mu(A_n)}\right)^{(p^+_{|A_n} - p^-_{|A_n})} \ \leq \ M_q\,,$$

for every positive entire n. These two facts allow to prove the following result.

Theorem 3.3 *([6]) For every $q \in R_{p(\cdot)}$, the space $L^{p(\cdot)}(\Omega)$ has a lattice complemented subspace isomorphic to l_q.*

A generalization of theorem above is the following: if $(p_n)_n$ is a sequence in $R_{p(\cdot)}$, then $L^{p(\cdot)}(\Omega)$ has a complemented isomorphic copy of the Nakano sequence space $l^{(p_n)}$.

We pass now to present some results on the boundedness of the averaging and orthonormal operators P_A and T_A associated to disjoint sequences $A = (A_n)$. We present some sufficient conditions on $A = (A_n)$ in order to the projections P_A and T_A be bounded.

Recall that given two real numbers sequence $(p_n)_n \subset [1,\infty)$ and $(w_n)_n \subset (0,\infty)$, the *weighted Nakano sequence space* $l^{(p_n)}(w_n)$ is the sequence space

$$l^{(p_n)}(w_n) = \{(x_n)_n \in \mathbb{R}^{\mathbb{N}} : \sum_{n=1}^{\infty} |\frac{x_n}{r}|^{p_n} w_n < \infty \text{ for some } r > 0\},$$

equipped with the Luxemburg norm

$$||(x_n)||_{l^{(p_n)}(w_n)} = \inf\{r > 0 : \sum_{n=1}^{\infty} |\frac{x_n}{r}|^{p_n} w_n \leq 1\}.$$

The following definition is useful.

Definition *Let $A = (A_n)_n$ be a sequence of disjoint measurable subsets and $p : \Omega \to [1,\infty)$ be a measurable function with $p^+ < \infty$. We say that the sequence A is $p(\cdot)$-regular if the associated weighted Nakano sequence spaces $l^{(p^+_{|A_n})}(\mu(A_n))$ and $l^{(p^-_{|A_n})}(\mu(A_n))$ coincide.*

Theorem 3.4 *([7]) If $A = (A_n)_n$ is a $p(\cdot)$-regular sequence of disjoint measurable sets in Ω then the averaging operator P_A and the orthogonal projection T_A are bounded.*

A function $p : \Omega \in \mathbb{R}^m \to \mathbb{R}$ is said to be *locally log-Hölder continuous* if there exists a constant $M > 0$ such that for $x, y \in \Omega$

$$|p(x) - p(y)| \leq \frac{M}{\log(e + \frac{1}{||x-y||})}.$$

These functions play an important role in the study of $L^{p(\cdot)}(\Omega)$ spaces and maximal operators (c.f. [1], [3]). Obviously every bounded α-Lipschitz function (for $0 < \alpha \leq 1$) is locally log-Hölder continuous.

If a measurable function $p : \Omega \subset \mathbb{R}^m \to [1,\infty)$ is locally *log-Hölder continuous*, then the projections T_A and P_A are always bounded.

In general the $p(\cdot)$-regularity condition of sequence $A = (A_n)_n$ is not a necessary condition for the operators P_A or T_A be bounded as next example shows.

Let $\Omega = (0,\infty)$, and the sequence $A = (\,(n-1,n)\,)_n$. Choose R_n, $S_n \subset A_n$ with $R_n \bigcap S_n = \emptyset$, $\mu(R_n) = w_n$, $\mu(S_n) = 1 - w_n$ and $\sum_{n=1}^{\infty} w_n = 1$. Define the measurable function $p(\cdot)$ by $p_{|R_n} = 2$ and $p_{|S_n} = 1$ for all $n \in \mathbb{N}$. Then $P_A = T_A$. Obviously A is not $p(\cdot)$-regular since $l_2 \neq l_1$, nevertheless the averaging projection P_A is bounded.

The following restriction on the kind of sequences $(A_n)_n$ allow to obtain necessary and sufficient conditions for the boundedness of P_A.

Definition *Let $A = (A_n)_n$ be a sequence of disjoint measurable sets in Ω with $p^+_{|A_n} - p^-_{|A_n} = \delta_n > 0$, for all $n \in \mathbb{N}$. We say that A is $p(\cdot)$-balanced if there exist* *sequences $(r_n)_n$ and $(s_n)_n$ of positive numbers in $(0,1)$ and two constants $r,k \in (0,1)$, with $0 < \max_n\{s_n + r_n\} < r < 1$, such that $\mu(S_n) > k\mu(A_n)$ and $\mu(R_n) > k\mu(A_n)$, where*

$$S_n = A_n \bigcap p^{-1}([p^-_{|A_n}, p^-_{|A_n} + s_n\delta_n])$$

and

$$R_n = A_n \bigcap p^{-1}([p^+_{|A_n} - r_n\delta_n, p^+_{|A_n}]),$$

for all $n \in \mathbb{N}$.

Theorem 3.5 *([7]) Let $A = (A_n)_n$ be a $p(\cdot)$-balanced sequence of disjoint measurable sets in Ω. Then A is $p(\cdot)$-regular if and only if the associated averaging operator P_A is bounded.*

The proof of this takes some ideas from [4].

Remark 1 *The averaging operators P_A are not always bounded. Indeed, whatever the essential range $R_{p(\cdot)}$ is not just one point there exists a disjoint measurable subset sequence $A = (A_n)_n$ in Ω such that the projection P_A is not bounded (see [7] Prop. 4.14).*

Some open questions are the following:

1.- Given $2 < q < \infty$, determine under which conditions on the exponent function $p(\cdot)$ the space $L^{p(\cdot)}$ contains a subspace isomorphic to L^q. (Note that this happens f.i. in the very special case of $p(\cdot)$ be equal to q on a non-null measurable set).

2.- Characterize when the spaces $L^{p(\cdot)}$ are subprojective and superprojective. (Recall that an space E is called subprojective is every subspace of E contains a further subspace which is complemented in E).

References

[1] D. Cruz-Uribe, A. Fiorenza: *Variable Lebesgue spaces. Fundations and Harmonic Analysis.* Birkhausser 2013.

[2] L. Diening: *Maximal function on Musielak-Orlicz spaces and generalized Lebesgue spaces.* Bull. Sci. Math. **129** (8) (2005) 657–700.

[3] L. Diening, P. Harjulehto, P. Hästö, M. Ružička: *Lebesgue and Sobolev spaces with variable exponents.* LNM 2017, Springer-Verlag 2011.

[4] D. E. Edmunds, A. Nekvinda: *Averaging operators on l^{p_n} and $L^{p(x)}$.* Math. Inequalities and Appl. **5** (2002) 235–246.

[5] S. Guerre-Delabrière: *Classical sequences in Banach spaces.* Marcel Dekker 1992.

[6] F. L. Hernández, C. Ruiz: *l_q-structure of variable exponent spaces.* J. Math. Anal. Appl. 389 (2012) 899–907.

[7] F. L. Hernández, C. Ruiz: *Averaging and orthogonal operators on variable exponent spaces* $L^{p(\cdot)}$. J. Math. Anal. Appl. 413 (2014) 139–153.

[8] H. Hudzik: *On some equivalent conditions in Musielak-Orlicz spaces.* Comment. Math. Prace Mat. **24** (1984) 57–64.

[9] A. Kaminska, B.Turet: *Type and cotype in Musielak-Orlicz spaces.* Lecture Notes London Math. Soc. 38 (1990) 165–180.

[10] O.Kovacik, J.Rakosnik: *On spaces* $L^{p(x)}$ *and* $W^{k,p(x)}$. Czechoslovak Math. Journal **41** (1991) 592–618.

[11] A. K. Lerner: *On some questions related to the maximal operator on variable Lp spaces.* Trans. AMS **362** (8) (2010) 4229–4242.

[12] J. Lindenstrauss, L. Tzafriri: *Classical Banach Spaces II.* Springer-Verlag 1979.

[13] J. Lukes, L. Pick, D. Pokorny, *On geometric properties of the spaces* $L^{p(x)}$. Rev. Mat. Complut. **24** (2011) 111–130.

[14] J. Musielak: *Orlicz Spaces and Modular Spaces.* LNM 1034, Springer 1983.

[15] V. Peirats, C. Ruiz: *On* l^p*-copies in Musielak-Orlicz sequence spaces.* Arch. Math. **58** (1992) 164–173.

[16] L. P. Poitevin, Y. Raynaud: *Ranges of positive contractive projections in Nakano Spaces.* Indag. Math. **19** (2008) 441–464.

[17] M. Ružička: *Electrorheological fluids: modeling and mathematical theory.* LNM 1748, Springer-Verlag 2000.

Carousel wild knots are ambient homogeneous

Gabriela HINOJOSA*, Cynthia VERJOVSKY MARCOTTE and Alberto VERJOVSKY†‡

Centro de Investigación en Ciencias. Instituto de Investigación en Ciencias Básicas y Aplicadas
Universidad Autónoma del Estado de Morelos
62209 Cuernavaca, Morelos, México
gabriela@uaem.mx

Mathematics Department
St. Edwards University
78704 Austin, TX, USA
cynthiavm@gmail.com

Instituto de Matemáticas
Universidad Nacional Autónoma de México
62209 Cuernavaca, Morelos, México
alberto@matcuer.unam.mx

Dedicated to José María Montesinos on his 70th birthday.

ABSTRACT

A carousel wild knot $K \subset \mathbb{S}^3$ is defined to be the intersection of a nested sequence of solid tori, where each solid torus is the union of a finite number of homeomorphic "beads", each with a knotted strand of beads of the next level down (satisfying some further conditions). In this paper we will prove that carousel wild knots are ambient homogeneous. We note a family of examples of carousel wild knots are ones that are dynamically defined as the limit set Λ of a Kleinian group Γ generated by reflections through 2-spheres that are boundaries of a necklace of pearls covering a knot K [10].

2010 Mathematics Subject Classification: 57M30, 57M45, 57Q45, 30F14.
Key words: Wild knots, Kleinian Groups.

*Author partially supported by project CONACyT (Mexico), CB-2009-129939.
†Authors partially supported by PAPIIT (UNAM) #IN103914.
‡We thank Aubin Arroyo for Figures 6 and 7.

1. Introduction

In 1920's Antoine ([2]) found the first example of a wild arc embedded in 3-space and in 1948 Artin and Fox constructed in [8] a family of wild arcs in 3-space called *Fox-Artin* wild arcs. In 1924 Alexander constructed in [1] embeddings of $\mathbb{S}^2$ in $\mathbb{S}^3$ which are not tame, called *wild embeddings*, such as Alexander's horned sphere. These examples can be generalized to produce wild embeddings of $\mathbb{S}^n$ in $\mathbb{S}^{n+1}$ and $\mathbb{S}^n$ in $\mathbb{S}^{n+2}$, for $n \geq 2$. Such examples called *higher-dimensional wild knots* can be obtained, for instance, by spinning or suspending lower dimensional wild knots. Examples of wild knots in higher dimensions which are limit sets of geometrically finite Kleinian groups can be found in [4].

Definition 1.1 *A topologically embedded n-sphere $K^n \subset \mathbb{S}^{n+2}$ is called a topological knot. A knot in the 3-sphere is tame if it has a polygonal representative in its ambient isotopy class. We say that a point $x \in K$ is locally flat or locally tame if there exists an open neighborhood U of x such that there is a homeomorphism of pairs: $(U, U \cap K) \sim (\mathrm{Int}(\mathbb{B}^{n+2}), \mathrm{Int}(\mathbb{B}^n))$, where $\mathbb{B}^n \subset \mathbb{R}^n$ is the unit closed n-ball. Otherwise, x is said to be a wild point of K. A knot K is locally flat or locally tame if all its points are locally flat. Otherwise, we say K is a wild knot.*

H. R. Bing proved that a locally tame knot in $\mathbb{S}^3$ is tame ([3]). This implies that there exists an extension of the embedding of the knot to an embedding of the solid torus $\mathbb{S}^1 \times \mathbb{B}^2$ into the 3-sphere.

Remark 1.1 *It can be shown that an embedding $g : \mathbb{S}^1 \to \mathbb{S}^3$ determines a wild knot K if and only if the fundamental group $\pi_1(\mathbb{S}^3 - K)$ is infinitely generated. In fact, if $x \in K$ is a wild point and $\mathcal{U}$ is any open neighborhood of x then $\pi_1(\mathcal{U} - K)$ is infinitely generated (compare [10] Lemma 3.2).*

We recall that a compact surface $\Sigma \subset \mathbb{S}^3$ embedded in the 3-sphere is tame if it has a closed neighborhood V homemorphic to $\Sigma \times [-1, 1]$ by a homemorphism $\Phi : \Sigma \times [-1, 1] \to V$ such that $\Phi(\Sigma \times \{0\}) = \Sigma$. An embedded solid torus in $\mathbb{S}^3$ is said to be *tame* if its boundary is tame. Two tame solid tori B_1 and B_2 in the 3-sphere are said to be *concentric* if $B_1 \subset \mathrm{Int}(B_2)$ and $B_2 - \mathrm{Int}(B_1)$ is homeomorphic to $\mathbb{T}^2 \times [0, 1]$, where $\mathbb{T}^2$ is the 2-torus.

C.H. Edwards proved the following theorem ([7] Corollary 1):

Theorem 1.1 *Let $\{B_n\}_1^\infty$ be a sequence of tame solid tori in $\mathbb{S}^3$ such that $B_{n+1} \subset \mathrm{Int}(B_n)$ for $n \geq 1$ and not two of the solid tori in $\{B_n\}_1^\infty$ are concentric. If $J = \bigcap_{n=1}^{\infty} B_n$ is a simple closed curve, then J is wildly embedded in $\mathbb{S}^3$.*

Definition 1.2 *We say that a knot $K \subset \mathbb{S}^{n+2}$ is ambient homogeneous if given two points p, $q \in K$, there exists a homeomorphism $\psi : \mathbb{S}^{n+2} \to \mathbb{S}^{n+2}$ such that $f(K) = K$ and $f(p) = q$.*

In general wild knots behave very differently from tame knots, as illustrated by Remark 1.2. For instance, the wild knot K given by Artin and Fox (see Figure 1), can not be ambient homogeneous [8] since K contains just one wild point p, then it is not possible to give a homeomorphism $\psi : \mathbb{S}^3 \to \mathbb{S}^3$ such that $f(K) = K$ and $f(p) = q$, $q \neq p$, because any homeomorphism sends wild points into wild points, since being locally tame is invariant under homeomorphism. In contrast it can be shown that any tame knot is ambient homogeneous (compare proof of Lemma 3.1).

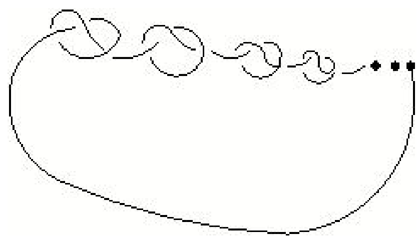


Figure 1: *The Artin-Fox wild knot: an example with one wild point.*

Examples of ambient homogeneous wild knots are given by H.G. Bothe in [5]. In fact, the author gives an example of a wild knot $\mathcal{K}$ in Euclidean space, which has the property that given any homeomorphism of $\mathcal{K}$ it can be extended to $\mathbb{R}^3$ (see also [6]).

In what follows, we will define carousel wild knots (Definition 2.1) and show that they are ambient homogeneous. In particular, we note dynamically defined wild knots are ambient homogeneous (see [9]), since the dynamical construction creates a carousel knot. In this case the associated nontrivial tangle types are all of a single kind and its mirror image.

2. Carousel Knots

Wild knots in $\mathbb{S}^3$ which are obtained as a nested intersection of solid tori are not necessarily ambient homogeneous since, for instance, it is easy to see that the Artin-Fox knot is the intersection of a sequence of nested solid tori. However if the sequence of tori satisfies the "carousel property" described below then we will show that the corresponding wild knot is indeed ambient homogeneous.

Definition 2.1 *A* carousel wild knot *is a wild knot $K \subset \mathbb{S}^3$ such that:*

1. *K is the nested intersection of a sequence of solid tori:*

$$K = \bigcap_{i\in\mathbb{N}} T_i,$$

where T_i is homeomorphic to $\mathbb{D}^2 \times \mathbb{S}^1$, where $\mathbb{D}^2$ is the closed 2-disk, and T_{i+1} is contained in the interior of T_i.

2. *Each T_i is a union of* beads *with disjoint interiors $T_i = \cup_{j=1,\ldots,j_i} B^i_j$ (the number of beads j_i depends on the level i), where bead B^i_j is homeomorphic to $\mathbb{D}^2 \times [0,1]$, and $B^i_j \cap B^i_{j+1} := D^i_j \cong \mathbb{D}^2_j \times \{1\} \cong \mathbb{D}^2_{j+1} \times \{0\}, \mod j_i$.*

3. *The beads are nested: For each level i, each B^i_j, $j \in \{1,\ldots,j_i\}$ contains a solid knotted strand (thickened one-strand* nontrivial *tangle) τ^i_j made up of $n_i > 1$ consecutive beads of the next level down: $B^{i+1}_{((j-1)\cdot n_1\cdot\dots\cdot n_i+1)}, \ldots, B^{i+1}_{(j\cdot n_1\cdot\dots\cdot n_i)}$. For a given level i the union of all nontrivial tangles τ^i_j is T_{i+1}. Note that $j_1 = n_1$ and $j_{i+1} = n_i \cdot j_i$ so $j_{i+1} = \prod_1^i n_i$.*

4. *For $j,\ k \in \{1,\ldots,j_i\}$ we have that (B^i_j, τ^i_j) and (B^i_k, τ^i_k) are homeomorphic pairs. In other words for each level i, τ^i_j and τ^i_k are tangle isomorphic.*

5. *If $m_i := \max_{r\in\{1,\ldots,j_i\}} \big(diameter(B^i_r)\big)$ then $\sum m_i < \infty$.*

Remark 2.1 *We could have n_i be different from bead to bead of the same level (notated as n^i_j for $j \in \{1,\ldots,j_i\}$, but for notational simplicity we fix the number of beads making up tangle τ^i_j, $j \in \{1,\ldots,j_i\}$ for each level i.*

Remark 2.2 *Since by construction K is the intersection of a nested sequence of* non-concentric *solid tori it follows from Theorem 1.3 that K is a wild knot. In fact since we will prove that K is homogeneous it will follow that K is wild* at every point.

Remark 2.3 *Clearly for every point $a \in K$, $\exists\{B^i_{r_i}\}_{i\in\mathbb{N}}$ such that $a \in B^i_{r_i}$. Since $\{B^i_{r_i}\}_{i\in\mathbb{N}}$ is a sequence of nested sets and* $\mathrm{diameter}(B^i_{r_i}) \to 0$ *by Cantor's Theorem $a = \cap B^i_{r_i}$. Therefore $\{r_1,\ldots,r_n,\ldots\}$ is an address for each $a \in K$. Note that there is an ambiguity in the address if a is on the intersecting bounding disks of two successive beads $B^n_{r_w}$ and $B^n_{r_w+1} \mod j_n$ (and hence of successive beads of all levels below that), but we can resolve the ambiguity by choosing, say, the label $r_w + 1 \mod j_n$ for each level where such ambiguities exist. Furthermore note that the set of points of K that are on the intersecting disks of successive beads of the same level is dense in K.*

Remark 2.4 *Wild knots defined as the limit set Λ of a Kleinian group Γ generated by reflections through 2-spheres that are boundaries of a necklace of pearls covering a knot K are carousel wild knots (see Section 4 for more details). We will call these knots* dynamically defined *wild knots.*

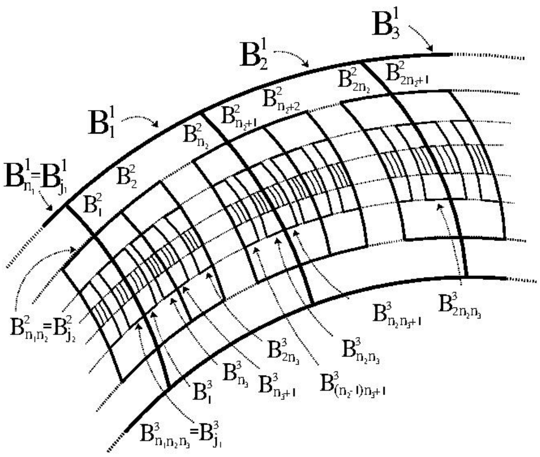


Figure 2: *Schematic diagram of the nesting of beads.*

3. Ambient Homogeneity

In this section we show that a wild carousel knot K is ambient homogeneous. In other words, given p, $q \in K$, we shall construct a homeomorphism $f : \mathbb{S}^3 \to \mathbb{S}^3$ such that $f(K) = K$ and $f(p) = q$.

Recall that K is the intersection of a nested decreasing sequence of solid tori T_i ($i \in \mathbb{N}$), *i.e.*, $K = \cap_{i=1}^{\infty} T_i$; such that $T_i \subset \mathrm{Int}(T_{i+1})$, and T_i is the union of j_i beads B^i_j whose diameters are smaller than $m_i > 0$ where $m_i \to 0$.

As noted in Remark 2.3 there exist two sequences of nested beads $\{B^i_{h_i}\}$, $\{B^i_{k_i}\}$, where $B^i_{h_i}$, $B^i_{k_i} \subset T_i$, such that $p = \cap_{i=1}^{\infty} B^i_{h_i}$ and $q = \cap_{i=1}^{\infty} B^i_{k_i}$. To construct f we will inductively construct a uniformly Cauchy sequence of homeomorphisms $\{f_l\}_{l\in\mathbb{N}}$ (which are the identity outside of a regular neighborhood of T_1), where each f_l maps beads of that level to beads of the same level, and in particular $f_i(B^i_{h_i}) = B^i_{k_i}$. Then $f := \lim f_i$ will be our desired homeomorphism.

Definition 3.1 *The* collar *of a solid torus T, $\mathscr{C}(T)$, is the closed regular neighborhood of a solid torus minus the interior of T, that is $N(T) - \mathrm{Int}(T) \cong \mathbb{T}^2 \times \mathbb{I}$ (where $\mathbb{I} \cong [0,1]$). The* outer boundary *of $\mathscr{C}(T)$ is $\partial\mathscr{C}(T) - \partial T$ and the* inner boundary *is the one that conicides with ∂T. If $T = T_i$ is made up of beads B^i_j for $j = 1, \ldots, j_i$ so $T = \cup_j B^i_j$ we define the* boundary annulus *of a bead B^i_j, $A^i_j := \partial T \cap B^i_j$ and the* collar *of a bead B^i_j to be $C^i_j \cong A^i_j \times \mathbb{I}$ so that $C^i_j \cap C^i_{j+1} = L^i_j \mod j_i$ is an annulus and $\mathscr{C}(T) = \cup_{j=1}^{j_i} C^i_j$.*

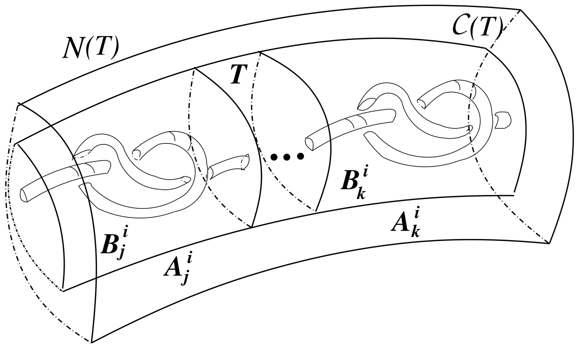


Figure 3: *Bead and collar.*

The first lemma will establish a technique to map the *collar* of a solid torus T to itself in such a way that it is the identity on one boundary and maps parallel annuli to parallel annuli (compare Figure 4). This will allow us to prove the next two lemmas. The first of these will establish the base of the induction, and the second will be the inductive step to construct our family of functions.

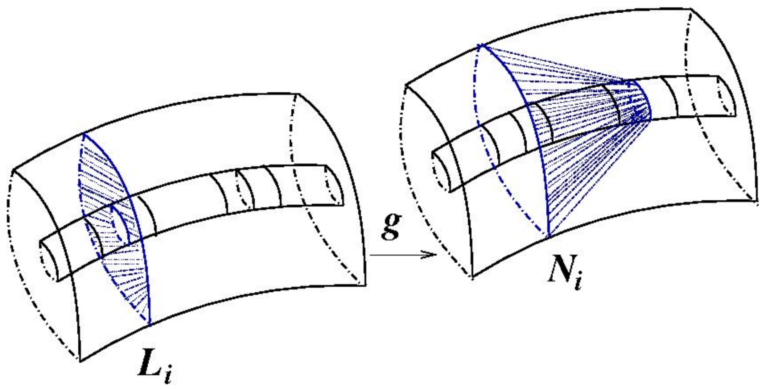


Figure 4: *Mapping the inner torus and collar.*

Lemma 3.1 *Let $\mathscr{C}(T_i)$ be the collar of a solid torus $T_i = \cup_j B^i_j$. Let L_j for $i \in \{1,\ldots,j_i\}$ be the boundary annuli of the beads. Then there exists a homeomorphism $g_i : \mathscr{C}(T_i) \to \mathscr{C}(T_i)$ that is the identity on the outer boundary torus of $\mathscr{C}(T)$ and maps A^i_j to $A^i_{j+r} \mod j_i$ for a given $r \in \{1,\ldots,j_1-1\}$.*

Proof. Recall $L_j = C^i_j \cap C^i_{j+1} \mod j_i$. Then we want to map L_j to an annulus N_j whose boundary is the outer boundary of L_j and its inner boundary is mapped to the inner boundary of L_{j+r}.

Since $\mathscr{C}(T_i)$ is a collar of T_i, there exists a homeomorphism $\psi_i : \mathbb{S}^1 \times \mathbb{D}^2_{v_i} \to \mathscr{C}(T_i)$ such that $\psi_i|_{\mathbb{S}^1\times\mathbb{D}^2_{r_i}} = \partial T_i$ $v_i > r_i > 0$.

According to Definition 2.1, we can subdivide $\mathbb{S}^1$ in j_i arcs of the same length (t_s, t_{s+1}) for $s = 1, \ldots, j_i \mod j_i$, such that up to reparametrization if necessary, $\psi_i((t_s, t_{s+1}) \times \mathbb{S}^1_{r_i}) = A^i_s$. In particular $\psi_i((t_j, t_{j+1}) \times \mathbb{S}^1_{r_i}) = A^i_j$ and $\psi_i((t_{j+r}, t_{j+r+1}) \times \mathbb{S}^1_{r_i}) = A^i_{j+r}$. Let $\alpha \in (0, 1)$ such that $e^{2\pi\imath\alpha} t_j = t_{j+r}$.

We define $g : \mathscr{C}(T_i) \to \mathscr{C}(T_i)$ as $g(x) = \psi_i(e^{2\pi\imath\alpha\frac{v_i-r}{v_i-r_i}} t, re^{2\pi\imath\theta})$, where $\psi_i^{-1}(x) = (t, re^{2\pi\imath\theta})$, $r_i \le r \le v_i$, $t \in \mathbb{S}^1$ and $\theta \in [0, 1]$ (see Figure 4). By construction g is a homeomorphism that is the identity on the outer boundary torus of $\mathscr{C}(T)$, this is $g|_{\partial\mathscr{C}(T_i)-\partial T_i} = \mathrm{id}$, and $g(A^i_j) = A^i_{j+r} \mod j_i$ for a given $r \in \{1, \ldots, j_1 - 1\}$. This proves our lemma. $\square$

Remark 3.1 *Consider the closed regular neighborhood $N(T_i)$ of T_i. Notice that from the previous proof we can extend g to a canonical homeomorphism $\tilde{g} : N(T_i) \to N(T_i)$ as follows:*

$$\tilde{g}(x) = \begin{cases} g(x) & \text{if } x \in \mathscr{C}(T_i), \\ \psi_i(e^{2\pi\imath\alpha} t, re^{2\pi\imath\theta}) & \text{if } x \in T_i. \end{cases}$$

where $\psi_i^{-1}(x) = (t, re^{2\pi\imath\theta})$, $0 \le r \le r_i$, $t \in \mathbb{S}^1$ and $\theta \in [0, 1]$. Since $g(x) = \psi_i(e^{2\pi\imath\alpha} t, re^{2\pi\imath\theta})$ for $x \in \partial T_i$, it follows that $\tilde{g}$ is a well-defined homeomorphism.

Lemma 3.2 *Let $p \ne q \in K$ and $V(T_1)$ be a closed regular neighborhood of T_1. Then there exists a homeomorphism $f_1 : \mathbb{S}^3 \to \mathbb{S}^3$ such that:*

1. $f_1(B^1_{h_1}) = B^1_{k_1}$.
2. $f_1(B^1_{h_1+s}) = B^1_{k_1+s} \mod j_1$.
3. *f_1 sends the boundary of each bead onto the corresponding boundary of its image; this is, $f_1(\partial B^1_{h_1+s}) = \partial B^1_{k_1+s} \mod j_1$.*
4. $f_1|\mathbb{S}^3 - \mathrm{Int}(V(T_1)) = \mathrm{id}$.
5. $d(\mathrm{id}, f_1) < \frac{j_1}{2} \cdot m_1$.

Proof. Let $V(T_1)$ be a closed regular neighborhood of T_1. Outside of $\mathrm{Int}(V(T_1))$ we will define f_1 to be the identity. In T_1 we will have f_1 send beads of the first level to beads of the same level, and in particular $B^1_{h_i}$ to $B^1_{k_i}$. Consider the *collar*

$\mathscr{C}(T_1)$ of T_1, that is $V(T_1) - \mathrm{Int}(T_1)$. By the previous Lemma there exists a homeomorphism $g_1 : \mathscr{C}(T_1) \to \mathscr{C}(T_1)$ that is the identity on the outer boundary torus of $\mathscr{C}(T)$ and maps $A^1_{h_1+r}$ to $A^1_{k_1+r} \mod j_1$ for a given $r \in \{0, 1, \ldots, j_1 - 1\}$. Moreover, by Remark 3.1 we can exted g_1 to a homeomorphism $\tilde{g_1} : V(T_1) \to V(T_1)$. Notice that $\tilde{g_1}|_{\partial V(T_1)} = \mathrm{id}$, hence there exists a homeomorphism $f_1 : \mathbb{S}^3 \to \mathbb{S}^3$ as follows: $f_1|_{V(T_1)} = \tilde{g_1}$ and $f_1(x) = x$ for $x \in \mathbb{S}^3 \setminus \mathrm{Int}(V(T_1))$.

By construction $f_1(B^1_{h_1+r}) = B^1_{k_1+r}$ and $f_1(\partial B^1_{h_1+r}) = \partial B^1_{k_1+r}$ for $r \in \{0, 1, \ldots, j_1 - 1\}$, and $d(\mathrm{id}, f_1) < \frac{j_1}{2} \cdot m_1$. This proves our Lemma. □

We have the base for our inductive construction; now we must show our inductive step.

Lemma 3.3 *Let $p \neq q \in K$ and let $f_i : \mathbb{S}^3 \to \mathbb{S}^3$ be a homeomorphism such that*

1. $f_i(B^i_{h_i}) = B^i_{k_i}$.
2. $f_i(B^i_{h_i+s}) = B^i_{k_i+s} \mod j_i$.
3. *f_i sends the boundary of each bead onto the corresponding boundary of its image; this is, $f_i(\partial B^i_{h_i+s}) = \partial B^i_{k_i+s} \mod j_i$.*
4. $f_i|\mathbb{S}^3 - \mathrm{Int}(V(T_i)) = \mathrm{id}$.

Then there exists a homeomorphism $f_{i+1} : \mathbb{S}^3 \to \mathbb{S}^3$ such that

1. *$f_{i+1}(x) = f_i(x) = x$ for any $x \in (\mathbb{S}^3 - \mathrm{Int}\, V(T_1))$,*
2. *$f_{i+1}(x) = f_i(x)$ for any $x \in (\mathbb{S}^3 - V(T_{i+1}))$, where $V(T_{i+1})$ is a regular neighborhood of T_{i+1} such that $V(T_{i+1}) \subset \mathrm{Int}(T_{i+1})$,*
3. $d(f_i, f_{i+1}) < 2 \cdot m_i$.

Proof. Let $V(T_{i+1})$ be a closed regular neighborhood of T_{i+1}. Outside of $\mathrm{Int}(V(T_{i+1}))$ we will define f_{i+1} to be f_i. In T_{i+1} we will have f_{i+1} send beads of the $(i+1)^{th}$-level to beads of the same level, and in particular $B^{i+1}_{h_i}$ to $B^{i+1}_{k_i}$. Consider the *collar* $\mathscr{C}(T_{i+1})$ of T_{i+1}, that is $V(T_{i+1}) - \mathrm{Int}(T_{i+1})$. By a slight modification of the argument used to prove Lemma 3.1 and Remark 3.1 there exists a homeomorphism $\widetilde{g_{i+1}} : V(T_{i+1}) \to V(T_{i+1})$ that coincedes with f_i on $\partial V(T_{i+1})$ and maps $A^{i+1}_{h_{i+1}+r}$ to $A^{i+1}1_{k_{i+1}+r} \mod j_{i+1}$ for a given $r \in \{0, 1, \ldots j_{i+1} - 1\}$. We define the map $f_{i+1} : \mathbb{S}^3 \to \mathbb{S}^3$ as follows: $f_{i+1}|_{V(K_1)} = \widetilde{g_{i+1}}$ and $f_{i+1}(x) = f_i(x)$ for $x \in \mathbb{S}^3 \setminus \mathrm{Int}(V(K_{i+1}))$. Observe that f_{i+1} is a homeomorphism, since $f_i(x)$ and $\widetilde{g_{i+1}}$ are homeomorphisms and $\widetilde{g_{i+1}}|_{\partial V(K_{i+1})} = f_i$ (see Figure 5).
By construction $f_{i+1}(B^{i+1}_{h_{i+1}+r}) = B^{i+1}_{k_{i+1}+r}$ and $f_{i+1}(\partial B^{i+1}_{h_{i+1}+r}) = \partial B^{i+1}_{k_{i+1}+r}$ for $r \in \{0, 1, \ldots, j_{i+1} - 1\}$. Notice that by Definition 2.1 and Remark 2.3 $d(B^{i+1}_{h_i}, B^{i+1}_{k_i}) <$

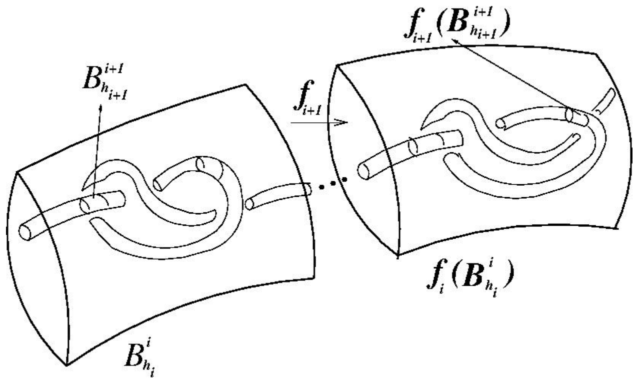


Figure 5: *Construction of* f_{i+1}.

$2 \max\limits_{r\in\{1,\ldots,j_i\}} \big(\text{diameter}(B^i_r)\big)$, hence $d(f_i, f_{i+1}) < 2\cdot m_i$. This proves our lemma. □

We now have inductively constructed a family of homeomorphisms so at the l-stage, we have a homeomorphism $f_l : \mathbb{S}^3 \to \mathbb{S}^3$ such that $f_l|_{\mathbb{S}^3 - \text{Int}(T_{l-1})} = f_n$ for all $n < l$ and it sends $f_{l-1}(B^l_{h_l})$ onto $B^l_{k_l}$; *i.e.*, $f_l(f_{l-1}(B^l_{h_l})) = B^l_{k_l}$ and $f_l(f_{l-1}(B^l_{h_l+s})) = B^l_{k_l+s}$ mod n_l. Again, this map also sends the boundary of each bead onto the boundary of its image, and $d(f_{l-1}, f_l) < 2\cdot m_{l-1}$. Recall that $\sum m_k < \infty$.

Lemma 3.4 *$\{f_l\}_{l\in\mathbb{N}}$ is a uniformly Cauchy sequence of continuous functions and therefore converges uniformly to a unique continuous function $f(x) = lim_{l\to\infty} f_l(x)$.*

Proof. Recall that $\sum m_i < \infty$ (see Definition 2.1). Given $\epsilon > 0$, let $N > 0$ such that $\sum_N^\infty m_i < \frac{\epsilon}{2}$. Then for all $x \in \mathbb{S}^3$ $d(f_n(x), f_m(x)) < 2\cdot\sum_n^m m_i < \epsilon$ for all $m > n > N$. □

Theorem 3.1 *Carousel knots are ambient homogeneous.*

Proof. First of all, note that by construction $f(p) = q$, since $d(f_i(p), q) < m_i$ and $\sum m_i < \infty$.

In the previous Lemma, we have constructed a sequence of homeomorphisms $\{f_l\}_{l\in\mathbb{N}}$ from $\mathbb{S}^3$ onto itself which is uniformly Cauchy and hence converges uniformly to

$f(x) := \lim_{l\to\infty} f_l(x)$, $x \in \mathbb{S}^3$, and therefore f is a continuous map.

Notice that for any $x \in (\mathbb{S}^3 - K)$, there exists $n \in \mathbb{N}$ such that $x \in T_n - \mathrm{Int}(T_{n+1})$, hence $f_s(x) = f_{n+1}(x)$ for $s > n+1$. This implies that $f(x) = f_{n+1}(x)$, hence $f|_{\mathbb{S}^3-K} : (\mathbb{S}^3 - K) \to (\mathbb{S}^3 - K)$ is bijective. Furthermore note that the set of points of K that are on the intersecting disks of successive beads of the same level is dense in K, so f_i becomes stationary at some point if both p and q are such points (see Remark 2.3).

Next we will verify that f is bijective on K. Let $a \neq b \in K$ with $a = \cap_{i=1}^{\infty} B^i_{r_i}$ and $b = \cap_{i=1}^{\infty} B^i_{t_i}$, where $B^i_{r_i}, B^i_{t_i} \subset T_i$. Since $a \neq b$ there exists $N \in \mathbb{N}$ such that $B^i_{r_i} \cap B^i_{t_i} = \emptyset$ for $i > N$. One has: $f(a) = \cap_{i=1}^{\infty} f_i(B^i_{r_i})$, and $f(b) = \cap_{i=1}^{\infty} f_i(B^i_{t_i})$ and $f_i(B^i_{r_i}) \cap f_i(B^i_{t_i}) = \emptyset$ for $i > N$, since f_i is a homeomorphism. Therefore $f(a) \neq f(b)$ so that f is injective.

Since $K = \cap_{k=1}^{\infty} T_k$, and $f(T_k) = T_k$ then $f(K) = K$. This implies that $f|_K : K \to K$ is surjective and hence it is a bijection. Summarizing $f : \mathbb{S}^3 \to \mathbb{S}^3$ is a bijective, continuous map from a compact set to itself, hence f^{-1} is continuous. Therefore f is a homeomorphism such that $f(K) = K$ and $f(p) = q$, and K is ambient homogeneous.

□

4. Dynamically defined wild knots

Consider a polygonal knot K_1 and let V_1 be a collection of n_1 consecutive tangent closed 3-balls $B^1_1, B^1_2, \ldots, B^1_{n_1}$ in $\mathbb{S}^3$ that cover K_1 as in Figure 6. This collection is called a *necklace* of n_1-pearls ($n_1 \geq 3$) subordinate to K_1. The boundary of each B^1_j is a 2-sphere denoted by Σ^1_j, $j = 1, \ldots, n_1$.

Figure 6: *A pearl-necklace subordinate to the trefoil knot.*

Let Γ be the group generated by reflections I_j, through Σ^1_j ($j = 1, \ldots, n_1$). Then Γ is a Kleinian group and its limit set Λ is a wild knot (compare [11] and [12]), which it is called a *dynamically defined wild knot*. This limit set has very interesting topological properties, for instance it is ambient homogeneous as it was shown in [9]. We will give a brief description of its construction, for more details see [10] and [9].

Notice that if we reflect with respect to each Σ^1_j, both a mirror image of K_1 and the corresponding necklace $V_1 - B^1_j$ are mapped into the ball B^1_j. This implies that B^1_j contains a solid knotted strand $\tau^1_j = I_j(V_1 - B^1_j)$. After reflecting with respect to each Σ^1_j, we obtain a new necklace V_2 of $j_2 = n_1(n_1 - 1)$ pearls, subordinate to a new knot K_2 which is isotopic to the connected sum of K_1 and n copies of its mirror image $\bar{K_1}$. Observe that in this case all the strands τ^1_j are the same two knot types, K_1 and $\bar{K_1}$, and $V_2 \subset V_1$ (see Figure 7).

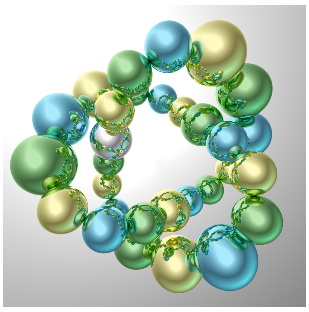


Figure 7: *A pearl-necklace after the first stage of the reflecting process.*

If we continue this process to the m-stage, we obtain a new necklace V_m consisting of $j_m = n_1(n_1 - 1)^{m-1}$ pearls, subordinate to a polygonal knot K_m. By construction, $V_m \subset V_{m-1}$ and all the strands τ^{m-1}_j are the same two knot types, K_1 and $\bar{K_1}$.

The limit set $\Lambda(\Gamma)$ given by

$$\Lambda(\Gamma) = \varprojlim_m V_m = \bigcap_{m=1}^{\infty} V_m$$

is a wild knot (see [11], [12]). Moreover from work done in [9] we have that Λ is a carousel knot. Next, we will explain briefly this fact (compare [10]).

Let T_1 be a closed tubular neighborhood of K_1 and $\pi_1 : T_1 \to K_1$ be the projection. We can assume that $\pi_1^{-1}(\{x\})$ is an euclidean 2-disk of radius $r_1 > 0$ independent of x. If $\{p^1_1, \ldots, p^1_{n_1}\}$ are the points of tangency of consecutive pearls of V_1, we can also assume that $\pi_1^{-1}(\{p^1_j\})$ is tangent to the consecutive pearls at p^1_j $(1 \leq j \leq n_1)$.

The tubular neighborhood T_1 is the union of n_1 "solid cylinders" $V_1^1, \dots, V_{n_1}^1$, where $V_j^1 = \pi_1^{-1}(\{[p_j^1, p_{j+1}^1]\})$ is called solid cylinder, since it is homeomorphic to a solid cylinder $C = \mathbb{D}^2 \times [0,1]$ (see Figure 8).

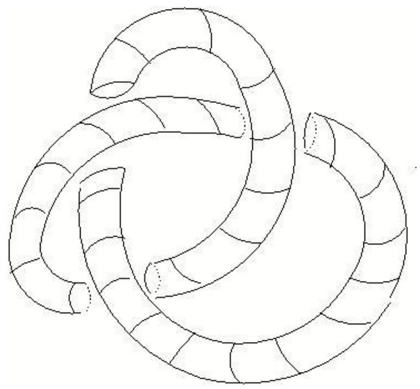


Figure 8: *A tubular neighbourhood as a union of "cylinders".*

For the second stage, we have a pearl necklace V_2 with j_2 pearls subordinate to the polygonal knot K_2. Let $T_2 \subset \mathrm{Int}(T_1)$ be a closed tubular neighborhood of K_2 and $\pi_2 : T_2 \to K_2$ be the projection. We again assume that $\pi_2^{-1}(\{x\})$ is an euclidean 2-disk of radius $r_2 > 0$ independent of x such that $r_2 < r_1$. Notice that the points $\{p_1^1, \dots, p_{n_1}^1\}$ are also points of tangency of consecutive pearls of V_2. We will denote by $\{p^2_{(i-1)(n_1-1)+1}, \dots, p^2_{(i-1)(n_1-1)+n_1-1}\} \subset V_2$ the corresponding points of tangency of consecutive pearls of $V_i^1 \cap T_2$, $1 \le i \le n_1$. We can again assume that $\pi_2^{-1}(\{p^2_{(i-1)(n_1-1)+k}\})$ is tangent to the consecutive pearls at $p^2_{(i-1)(n_1-1)+k}$ $(1 \le i, k \le n_1 - 1)$. The tubular neighborhood T_2 is the union of j_2 solid cylinders $V^2_{(i-1)(n_1-1)+k}$, where $V^2_{(i-1)(n_1-1)+k} = \pi_2^{-1}(\{[p^2_{(i-1)(n_1-1)+k}, p^2_{(i-1)(n_1-1)+k+1}]\})$. Note that each solid cylinder V_i^1 contains a solid knotted strand τ_i^1 made up of $n_1 - 1$ consecutive beads $V^2_{(i-1)(n_1-1)+1}, \dots, V^2_{(i-1)(n_1-1)+(n_1-1)}$ and all the strands τ_j^i are the same two knot types, K_1 and $\bar{K_1}$, see Figure 9.

We continue inductively, so at the end of the m-stage of the reflecting process, we have the pearl necklace V_m consisting of j_m pearls subordinate to the polygonal knot K_m. Let T_m be a closed tubular neighborhood of K_m such that $T_m \subset \mathrm{Int}(T_{m-1})$. Let $\pi_m : T_m \to K_m$ be the projection. We assume that $\pi_m^{-1}(\{x\})$ is an euclidean 2-disk of radius $r_m > 0$ independent of x such that $r_m < r_{m-1}$. We again denote the corresponding points of tangency of consecutive pearls of $V_i^{m-1} \cap T_m$, $1 \le i \le n_1$, by $\{p^m_{(i-1)(n_1-1)^{m-1}+1}, \dots, p^m_{(i-1)(n_1-1)^{m-1}+n_1-1}\} \subset V_m$. We can again assume that $\pi_m^{-1}(\{p^m_{(i-1)(n_1-1)^{m-1}+k}\})$ is tangent to the consecutive pearls at $p^m_{(i-1)(n_1-1)^{m-1}+k}$

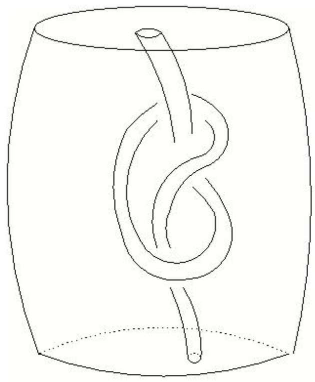


Figure 9: *A solid knotted strand.*

$(1 \le i,\, k \le n_1 - 1)$. The tubular neighborhood T_m is the union of j_m solid cylinders $V^m_{(i-1)(n_1-1)^{m-1}+k}$, where

$$V^m_{(i-1)(n_1-1)^{m-1}+k} = \pi_m^{-1}(\{[p^m_{(i-1)(n_1-1)^{m-1}+k}, p^m_{(i-1)(n_1-1)^{m-1}+k+1}]\}).$$

Note that each solid cylinder V_i^{m-1} contains a solid knotted strand τ_i^{m-1} made up of $n_1 - 1$ consecutive beads $V^m_{(i-1)(n_1-1)^{m-1}+1}, \ldots, V^m_{(i-1)(n_1-1)^{m-1}+(n_1-1)}$.

By construction $\lim_{k\to\infty} r_k = 0$ and $\Lambda = \cap_{k=1}^{\infty} T_k$. Notice that Λ satisfies all the conditions of Definition 2.1, this implies that Λ *is a carousel knot.*

References

[1] J. W. Alexander: *An example of a simply connected surface bounding a region which is not simply connected.* Proc. National Academy of Sciences USA **10**(1) (1924) 8–10.

[2] L. Antoine: *Sur l'homéeomorphie de deux*
gures et de leurs voisinages. J. Mathématiques Pures et Appliquées **86** (1921) 221–235.

[3] H. R. Bing: *Locally tame sets are tame.* Ann.Math. Second Series, **59**(1) (1948) 145–158.

[4] M. Boege, G. Hinojosa, A. Verjovsky: *Wild knots in higher dimensions as limit sets of Kleinian groups.* Conformal Geometry and Dynamics AMS **13** (2009) 197–216.

[5] H. G. Bothe: *Ein homogen wilder Knoten.* Fundamenta Mathematicae **850** (1967) 271–283.

[6] H. G. Bothe: *Homogeneously wild curves and infinite knot products.* Fundamenta Mathematicae **113**(2) (1981) 91–111.

[7] C. H. Edwards, Jr.: *Concentric tori and tame curves in* $\mathbb{S}^3$. In *Topology of 3-Manifolds and Related Topics, Proceedings of the University of Georgia Institute*(M. K. Fort, Jr., ed.). Prentice-Hall 1962.

[8] R. H. Fox, E. Artin: *Some wild cells and spheres in three-dimensional space.* Ann. Math., Second Series **49**(4) (1948) 979–990.

[9] G. Hinojosa: *Wild knots as limit sets of Kleinian groups.* Contemporary Mathematics **389** (2005) 125–139.

[10] G. Hinojosa, A. Verjovsky: *Homogeneity of dynamically defined wild knots.* Revista Matemática Complutense **19**(1) (2006) 101–111.

[11] M. Kapovich: *Topological aspects of Kleinian groups in several dimensions.* MSRI Preprint 1992. Updated in 2002, submitted to Proceedings of the 3rd Ahlfors-Bers Colloquium.

[12] B. Maskit: *Kleinian Groups.* Springer-Verlag 1987.

Global bifurcation for Fredholm operators

Julián López-Gómez*

Department of Applied Mathematics
Complutense University of Madrid
28040 Madrid, Spain
julian@mat.ucm.es

Al profesor J. M. Montesinos Amilibia, con ocasión de su jubilación, con respeto y admiración, en recuerdo de un magnífico curso de teoría de nudos que me impartió en Jarandilla de la Vera durante mi periodo de formación pre-doctoral. Nunca olvidaré su gran destreza matemática y su contagioso entusiasmo. Ni, por supuesto, su sabia conversación con mi entrañable maestro Miguel de Guzmán Ozámiz durante una comida conjunta que disfrutamos en el Parador de Oropesa de regreso a Madrid. Me fue concedida la gracia de aprender todo lo que sé por boca de los mejores maestros.

ABSTRACT

In this note we review and update a global bifurcation theorem for Fredholm operators of class $\mathcal{C}^1$ attributable to J. López-Gómez and C. Mora-Corral [17] and derive from it a global version of the celebrated local theorem of M. G. Crandall and P. H. Rabinowitz [5] on bifurcation from simple eigenvalues.

2010 Mathematics Subject Classification: 47J15, 58C40.
Key words: Bifurcation theory, Fredholm operators, topological degree, algebraic multiplicities, compact components.

1. Introduction

Throughout this paper, given two real Banach spaces, U and V, we will denote by $\mathcal{L}(U,V)$ the space of the bounded linear operators from U to V, and by $\mathrm{Fred}_0(U,V)$ the subset of $\mathcal{L}(U,V)$ consisting of all Fredholm operators of index zero. Subsequently, for a given $L \in \mathcal{L}(U,V)$, we will denote by $N[L]$ and $R[L]$ the 'Null space', or kernel, of L,

*Supported by the Project MTM2012-30669 of the Ministry of Economy and Competitiveness of Spain and the Institute of Interdisciplinary Mathematics of Complutense University of Madrid.

and the 'range', or image, of L, respectively. We recall that an operator $L \in \mathcal{L}(U, V)$ is said to be a Fredlhom operator if

$$\dim N[L] < \infty \quad \text{and} \quad \operatorname{codim} R[L] < \infty.$$

In such case, $R[L]$ must be closed, and the index of L is defined by

$$\operatorname{ind}[L] := \dim N[L] - \operatorname{codim} R[L].$$

Thus, $L \in \mathrm{Fred}_0(U, V)$ if

$$\dim N[L] = \operatorname{codim} R[L] < \infty.$$

Naturally, if $\mathrm{Fred}_0(U, V) \neq \emptyset$, then U and V must be isomorphic. So, it would not be a serious restriction assuming that $U = V$. In such case, we simply denote

$$\mathrm{Fred}_0(U) := \mathrm{Fred}_0(U, U).$$

The most paradigmatic class of functions in $\mathrm{Fred}_0(U)$ are the compact perturbations of the identity map, I_U. An operator $T \in \mathcal{L}(U, V)$ is said to be compact if the closure $\overline{T(B)}$ is a compact subset of V for all bounded subset $B \subset U$. In this paper, we will denote by $\mathcal{K}(U, V)$ the subset of $\mathcal{L}(U, V)$ of all compact operators. Another important subset of $\mathcal{L}(U, V)$ is the set of all isomorphism from U to V, $\mathrm{Iso}(U, V)$. Naturally, we will denote

$$\mathcal{L}(U) := \mathcal{L}(U, U), \qquad \mathcal{K}(U) := \mathcal{K}(U, U), \qquad \mathrm{Iso}(U) := \mathrm{Iso}(U, U).$$

The main goal of this paper consists in analyzing the structure of the components of $\mathfrak{F}^{-1}(0)$ from the point of view of global bifurcation theory, where

$$\mathfrak{F} \; : \; \mathbb{R} \times U \to V \tag{1.1}$$

is a general continuous map satisfying the following requirements:

(F1) For each $\lambda \in \mathbb{R}$, the partial map $\mathfrak{F}(\lambda, \cdot)$ is of class $\mathcal{C}^1(U, V)$ and

$$D_u\mathfrak{F}(\lambda, u) \in \mathrm{Fred}_0(U, V) \qquad \text{for all} \;\; u \in U, \tag{1.2}$$

where D_u stands for differentiation with respect to $u \in U$.

(F2) $D_u\mathfrak{F} : \mathbb{R} \times U \to \mathcal{L}(U, V)$ is a continuous map.

(A3) There exists $\theta \in \mathcal{C}(\mathbb{R}, U)$ such that $\mathfrak{F}(\lambda, \theta(\lambda)) = 0$ for all $\lambda \in \mathbb{R}$.

By performing the change of variable

$$\mathfrak{G}(\lambda, u) := \mathfrak{F}(\lambda, u + \theta(\lambda)), \qquad (\lambda, u) \in \mathbb{R} \times U,$$

and inter-exchanging $\mathfrak{F}$ by $\mathfrak{G}$, we can assume, instead of (A3), (F3) $\mathfrak{F}(\lambda, 0) = 0$ for all $\lambda \in \mathbb{R}$.

As $(\lambda, 0)$ is a given (known) zero, it will be referred to as the *trivial state*. The main goal of this paper is to update a very sharp property of the compact components of the set of nontrivial solutions of

$$\mathfrak{F}(\lambda, u) = 0, \qquad (\lambda, u) \in \mathbb{R} \times U, \tag{1.3}$$

bifurcating from the trivial solution $(\lambda, 0)$, which is a direct consequence of the main theorem of J. López-Gómez and C. Mora-Corral [17]. Then, we will adapt it to get a global sharp version of the celebrated local theorem of M. G. Crandall and P. H. Rabinowitz [5] on bifurcation from simple eigenvalues, whose number of applications is huge.

For any given $\lambda_0 \in \mathbb{R}$, it is said that $(\lambda_0, 0)$ is a *bifurcation point* of $\mathfrak{F} = 0$ from $(\lambda, 0)$ if

$$(\lambda_0, 0) \in \mathcal{S} := \text{closure}\left(\mathfrak{F}^{-1}(0) \cap [\mathbb{R} \times (U \setminus \{0\})]\right). \tag{1.4}$$

The set $\mathcal{S}$ will be referred to as the set of *nontrivial solutions* of $\mathfrak{F} = 0$. It consists of the *non-trivial solutions*

$$(\lambda, u) \in \mathfrak{F}^{-1}(0) \cap [\mathbb{R} \times (U \setminus \{0\})]$$

plus all possible bifurcation points from $(\lambda, 0)$. By a *component* of $\mathcal{S}$ it is meant a closed and connected subset of $\mathcal{S}$ which is maximal for the inclusion. So, by a component we really mean a connected component.

The distribution of this paper is as follows. Section 2 contains some basic preliminaries. Section 3 gives the concept of orientation and degree introduced by P. Benevieri and M. Furi. Section 4 collects the most relevant concepts and results of the theory of algebraic multiplicities, as they detect any change of orientation and hence, any global bifurcation phenomenon. Section 5 discusses the global bifurcation theorem and, finally, Section 6 derives from the previous results the global version of the local theorem of M. G. Crandall and P. H. Rabinowitz [5].

2. Preliminaries

Throughout this paper, for any map $\mathfrak{F}$ satisfying (F1), (F2) and (F3), we denote by

$$\mathfrak{L}(\lambda) := D_u\mathfrak{F}(\lambda, 0), \qquad \lambda \in \mathbb{R}, \tag{2.1}$$

the linearization of $\mathfrak{F}$ at the trivial state $(\lambda, 0)$. By (F2), $\mathfrak{L} \in \mathcal{C}(\mathbb{R}, \mathcal{L}(U, V))$. Moreover, since $\mathfrak{L}(\lambda) \in \text{Fred}_0(U, V)$,

$$\mathfrak{L}(\lambda) \in \text{Iso}\,(U, V) \quad \text{if} \quad \dim N[\mathfrak{L}(\lambda)] = 0.$$

Consequently, the *spectrum* of $\mathfrak{L}$ can be defined as

$$\Sigma := \Sigma(\mathfrak{L}) \equiv \{\lambda \in \mathbb{R} \;:\; \dim N[\mathfrak{L}(\lambda)] \geq 1\}. \tag{2.2}$$

Naturally, the resolvent set of $\mathfrak{L}$ is defined by $\varrho(\mathfrak{L}) := \mathbb{R} \setminus \Sigma$. Since $\mathfrak{L} \in \mathcal{C}(\mathbb{R}, \mathcal{L}(U,V))$ and $\mathrm{Iso}\,(U,V)$ is an open subset of $\mathcal{L}(U,V)$, $\varrho(\mathfrak{L})$ is open, possibly empty. Hence, $\Sigma(\mathfrak{L})$ is closed. Moreover, the next result holds.

Lemma 2.1 *Let $\lambda_0 \in \mathbb{R}$ be such that $(\lambda_0, 0)$ is a bifurcation point of $\mathfrak{F} = 0$ from $(\lambda, 0)$. Then, $\lambda_0 \in \Sigma(\mathfrak{L})$.*

Proof. Let (λ_n, u_n), $n \geq 1$, a sequence of $\mathbb{R} \times U$ with $u_n \neq 0$, $n \geq 1$, such that

$$\lim_{n\to\infty} (\lambda_n, u_n) = (\lambda_0, 0) \quad \text{and} \quad \mathfrak{F}(\lambda_n, u_n) = 0 \quad \text{for all } n \geq 1. \tag{2.3}$$

Then, setting

$$\mathfrak{N}(\lambda, u) := \mathfrak{F}(\lambda, u) - \mathfrak{L}(\lambda)u, \qquad (\lambda, u) \in \mathbb{R} \times U, \tag{2.4}$$

we have that

$$0 = \mathfrak{F}(\lambda_n, u_n) = \mathfrak{L}(\lambda_n)u_n + \mathfrak{N}(\lambda_n, u_n), \qquad n \geq 1. \tag{2.5}$$

Note that, thanks to (F3) and (2.1), we also have that

$$\mathfrak{N}(\lambda, 0) = 0, \qquad D_u\mathfrak{N}(\lambda, 0) = 0, \qquad \lambda \in \mathbb{R}. \tag{2.6}$$

Suppose $\lambda_0 \in \varrho(\mathfrak{L})$. Then, since (2.5) can be re-written as

$$\mathfrak{L}(\lambda_0)u_n = [\mathfrak{L}(\lambda_0) - \mathfrak{L}(\lambda_n)]u_n - \mathfrak{N}(\lambda_n, u_n) = 0, \qquad n \geq 1,$$

and $\mathfrak{L}(\lambda_0) \in \mathrm{Iso}\,(U,V)$, we find that

$$u_n = \mathfrak{L}^{-1}(\lambda_0)[\mathfrak{L}(\lambda_0) - \mathfrak{L}(\lambda_n)]u_n - \mathfrak{L}^{-1}(\lambda_0)\mathfrak{N}(\lambda_n, u_n), \qquad n \geq 1.$$

Hence, dividing by $\|u_n\|$ and taking norms yields

$$1 \leq \|\mathfrak{L}^{-1}(\lambda_0)\| \|\mathfrak{L}(\lambda_0) - \mathfrak{L}(\lambda_n)\| + \|\mathfrak{L}^{-1}(\lambda_0)\| \frac{\|\mathfrak{N}(\lambda_n, u_n)\|}{\|u_n\|}, \qquad n \geq 1. \tag{2.7}$$

By the continuity of $\mathfrak{L}(\lambda)$, (2.3) implies that

$$\lim_{n\to\infty} \|\mathfrak{L}(\lambda_0) - \mathfrak{L}(\lambda_n)\| = 0.$$

Moreover, according to (2.6),

$$\mathfrak{N}(\lambda_n, u_n) = \mathfrak{N}(\lambda_n, u_n) - \mathfrak{N}(\lambda_n, 0) = \int_0^1 D_u\mathfrak{N}(\lambda_n, tu_n)u_n\, dt$$

and hence,

$$\|\mathfrak{N}(\lambda_n, u_n)\| \leq \int_0^1 \|D_u\mathfrak{N}(\lambda_n, tu_n)\|\, dt\, \|u_n\|$$

for all $n \geq 1$. Thus, again by (2.6),

$$\limsup_{n\to\infty} \frac{\|\mathfrak{N}(\lambda_n, u_n)\|}{\|u_n\|} \leq \limsup_{n\to\infty} \int_0^1 \|D_u\mathfrak{N}(\lambda_n, tu_n)\|\, dt = 0.$$

Therefore, letting $n \to \infty$ in (2.7) yields $1 \leq 0$, which is impossible. This contradiction shows that $\lambda_0 \in \Sigma$ and ends the proof. $\square$

3. Orientation and degree for Fredholm maps

This section collects the concepts of orientation and topological degree for Fredholm maps of class $\mathcal{C}^1$ introduced by P. Benevieri and M. Furi [1]-[3], as well as some related findings by J. López-Gómez and C. Mora-Corral [17]. These concepts sharpen those derived from the parity introduced by P. M. Fitzpatrick and J. Pejsachowicz [9].

Subsequently, given three real Banach spaces, U, V and W, and an operator $L \in \mathrm{Fred}_0(U,V)$, we will denote by $\mathcal{F}(L)$ the (non-empty) set of finite-rank operators $F \in \mathcal{L}(U,V)$ such that $L+F \in \mathrm{Iso}\,(U,V)$. An equivalence relation can be defined in $\mathcal{F}(L)$ by declaring that $F_1, F_2 \in \mathcal{F}(L)$ are equivalent, $F_1 \sim_L F_2$, if

$$\det\,[(L+F_1)^{-1}(L+F_2)] > 0.$$

Since

$$(L+F_1)^{-1}(L+F_2) = I_U + (L+F_1)^{-1}(F_2 - F_1)$$

is a finite rank perturbation of the identity map I_U, its determinant can be defined as, e.g., in Section III.4.3 of T. Kato [11]. This relation possesses two equivalence classes. Each of them is called an *orientation* of L, and L is said to be oriented when an orientation has been chosen. In such case, that orientation is denoted by $\mathcal{F}_+(L)$ and we set $\mathcal{F}_-(L) := \mathcal{F}(L) \setminus \mathcal{F}_+(L)$.

Given two oriented operators, $L_1 \in \mathrm{Fred}_0(U,V)$ and $L_2 \in \mathrm{Fred}_0(V,W)$, their *oriented composition* is the operator L_2L_1 equipped with the orientation, $\mathcal{F}_+(L_2L_1)$, containing the operator $L_2F_1 + F_2F_1 + F_2L_1$, where $F_1 \in \mathcal{F}_+(L_1)$ and $F_2 \in \mathcal{F}_+(L_2)$. It is well defined in the sense that it does not depend on the choice of F_1 and F_2.

Let $L \in \mathrm{Iso}\,(U,V)$ be oriented. Its sign, $\mathrm{sgn} L$, is then defined by

$$\mathrm{sgn}\, L := \begin{cases} 1, & \text{if } 0 \in \mathcal{F}_+(L), \\ -1, & \text{if } 0 \in \mathcal{F}_-(L). \end{cases}$$

Te next result provides us with a very useful property of the sign. It is Lemma 2.1 of J. López-Gómez and C. Mora-Corral [17].

Lemma 3.1 *Let $L_1 \in \mathrm{Iso}\,(U,V)$ and $L_2 \in \mathrm{Iso}\,(V,W)$ be two oriented isomorphisms, and consider the oriented composition L_2L_1. Then,*

$$\mathrm{sgn}\,(L_2L_1) = \mathrm{sgn}\, L_2 \cdot \mathrm{sgn}\, L_1.$$

Next, we suppose X is a topological space and $\mathfrak{L} \in \mathcal{C}(X, \mathcal{L}(U,V))$ satisfies $\mathfrak{L}(x) \in \mathrm{Fred}_0(U,V)$ for all $x \in X$. An orientation of $\mathfrak{L}$ is a map $X \ni x \mapsto \alpha(x)$ such that $\alpha(x)$ is an orientation of $\mathfrak{L}(x)$ for all $x \in X$, and the map α satisfies the next continuity condition: for each $x_0 \in X$ and $F \in \alpha(x_0)$, there exists a neighborhood, $\mathcal{U}$, of x_0 in X such that $F \in \alpha(x)$ for all $x \in \mathcal{U}$. Although not every $\mathfrak{L}$ admits an orientation, the next result holds (see [1]-[3]).

Proposition 3.1 *Suppose X is a simply connected topological space. Then, every map $\mathfrak{L} \in \mathcal{C}(X, \mathcal{L}(U,V))$ such that $\mathfrak{L}(x) \in \mathrm{Fred}_0(U,V)$ for all $x \in X$ admits two orientations, $\mathcal{F}_+(\mathfrak{L})$ and $\mathcal{F}_-(\mathfrak{L})$, and each of them is uniquely determined by the orientation of $\mathfrak{L}(x)$ for an arbitrary $x \in X$.*

In this paper X is simply connected because X will be typically $\mathbb{R}$. Whenever X is simply connected and $\mathfrak{L} \in \mathcal{C}(X, \mathcal{L}(U,V))$ satisfies $\mathfrak{L}(x) \in \mathrm{Fred}_0(U,V)$ for all $x \in X$, we will think of $\mathfrak{L}$ as oriented by $\mathcal{F}_+(\mathfrak{L})$. Also, if $g \in \mathcal{C}^1(U,V)$ satisfies $Dg(x) \in \mathrm{Fred}_0(U,V)$ for all $x \in U$, then we will suppose that g is oriented, which means that an orientation, $\mathcal{F}_+(Dg)$, has been chosen for Dg. Similarly, any operator $\mathfrak{F} \in \mathcal{C}(\mathbb{R} \times U, V)$ satisfying (F1) and (F2) is assumed to be oriented by choosing an orientation, $\mathcal{F}_+(D_u\mathfrak{F})$, for $D_u\mathfrak{F}$.

Finally, we denote by $\mathcal{A}$ the set of (admissible) pairs, $(g, \mathcal{U})$, formed by an oriented function $g \in \mathcal{C}^1(U,V)$ such that $Dg(x) \in \mathrm{Fred}_0(U,V)$ for all $x \in U$, and an open subset $\mathcal{U} \subset U$ such that $g^{-1}(0) \cap \mathcal{U}$ is compact. According to P. Benevieri and M. Furi [1], a $\mathbb{Z}$-values degree is defined in $\mathcal{A}$, and it satisfies the same fundamental properties as the Leray-Schauder degree. Among them, the normalization, the additivity and the homotopy-invariance.

4. The algebraic multiplicity for Fredholm maps

Subsequently, given an open subinterval, J, of $\mathbb{R}$ and $r \in \mathbb{N} \cup \{\infty, \omega\}$, we will denote by $\mathcal{C}^r(J, \mathrm{Fred}_0(U,V))$ the set of maps of class $\mathcal{C}^r$ from J to $\mathcal{L}(U,V)$ with values in $\mathrm{Fred}_0(U,V)$; as usual, $\mathcal{C}^\omega$ stands for the set of real analytic maps. The next concept, introduced in [13, Def. 4.3.1], plays a pivotal role in the theory of algebraic multiplicities.

Definition 4.1 *Suppose J is an open subinterval of $\mathbb{R}$, $\mathfrak{L} \in \mathcal{C}^r(J, \mathrm{Fred}_0(U,V))$ for some integer $r \geq 1$, and $\lambda_0 \in J$; λ_0 is said to be an algebraic eigenvalue of $\mathfrak{L}$ if*

$$\dim N[\mathfrak{L}(\lambda_0)] \geq 1$$

and there are $C, \delta > 0$ and an integer $1 \leq k \leq r$ such that $\mathfrak{L}(\lambda) \in \mathrm{Iso}\,(U,V)$ for $0 < |\lambda - \lambda_0| < \delta$ and

$$\|\mathfrak{L}^{-1}(\lambda)\|_{\mathcal{L}(V,U)} \leq \frac{C}{|\lambda - \lambda_0|^k}, \qquad 0 < |\lambda - \lambda_0| < \delta; \tag{4.1}$$

λ_0 is said to be a k-algebraic eigenvalue of $\mathfrak{L}$ if, in addition, k is minimal.

The next result is a direct consequence from Theorems 4.4.1 and 4.4.4 of [13]. In many applications the dependence of $\mathfrak{L}(\lambda)$ in λ is real analytic.

Theorem 4.1 *Suppose J is an open subinterval of $\mathbb{R}$, $\mathfrak{L} \in \mathcal{C}^\omega(J, \mathrm{Fred}_0(U,V))$, and*

$$\Sigma := \{\lambda \in J \ : \ \dim N[\mathfrak{L}(\lambda)] \geq 1\}.$$

Then, either $\Sigma = J$, or Σ is a discrete subset of J. Moreover, if it is discrete, any $\lambda_0 \in \Sigma$ must be an algebraic eigenvalue of $\mathfrak{L}(\lambda)$, as discussed by Definition 4.1.

Actually, the complex counterpart of Theorem 4.1 also holds (see Chapter 8 of J. López-Gómez and C. Mora-Corral [18]). In the context of the Riesz-Schauder theory, $U = V$ and $\mathfrak{L}$ is given by

$$\mathfrak{L}(\zeta) = I_U - \zeta T, \qquad \zeta \in \mathbb{C},$$

with $T \in \mathcal{K}(U)$. As Theorem 4.1 is still valid in a complex setting and $\mathfrak{L}(0) = I_U$ is an isomorphism, it follows from Theorem 4.1 that $\Sigma(\mathfrak{L})$ is a discrete subset of $\mathbb{C}$. Moreover, any characteristic value of T must be a pole of the resolvent operator $(I_U - \zeta T)^{-1}$.

The next concept was introduced by J. Esquinas and J. López-Gómez [7] to extend the theorem of M. G. Crandall and P. H. Rabinowitz [5] on bifurcation from simple eigenvalues. Subsequently, for any given $\mathfrak{L} \in \mathcal{C}^r(J, \mathcal{L}(U,V))$ and $\lambda_0 \in J$ we will denote

$$\mathfrak{L}_0 := \mathfrak{L}(\lambda_0), \quad \mathfrak{L}_j = \frac{1}{j!}\frac{d^j \mathfrak{L}}{d\lambda^j}(\lambda_0), \qquad 1 \leq j \leq r.$$

Definition 4.2 *Suppose J is an open subinterval of $\mathbb{R}$, $\mathfrak{L} \in \mathcal{C}^r(J, \mathrm{Fred}_0(U,V))$ for some integer $r \geq 1$, and $\dim N[\mathfrak{L}(\lambda_0)] \geq 1$ for some $\lambda_0 \in J$. Then, given $1 \leq k \leq r$, λ_0 is said to be a k-transversal eigenvalue of $\mathfrak{L}(\lambda)$ if*

$$\bigoplus_{j=1}^{k} \mathfrak{L}_j(N[\mathfrak{L}_0] \cap \cdots \cap N[\mathfrak{L}_{j-1}]) \oplus R[\mathfrak{L}_0] = V \tag{4.2}$$

with

$$\dim \mathfrak{L}_k(N[\mathfrak{L}_0] \cap \cdots \cap N[\mathfrak{L}_{k-1}]) \geq 1.$$

The integer $k \geq 1$ is refereed to as the order of transversality of λ_0.

In such case, the algebraic multiplicity of $\mathfrak{L}(\lambda)$ at λ_0, $\chi[\mathfrak{L}; \lambda_0]$, is defined by

$$\chi[\mathfrak{L}; \lambda_0] := \sum_{j=1}^{k} j \cdot \dim \mathfrak{L}_j(N[\mathfrak{L}_0] \cap \cdots \cap N[\mathfrak{L}_{j-1}]). \tag{4.3}$$

Naturally, if $r = 1$, the transversality condition of M. G. Crandall and P. H. Rabinowitz [5] holds if, and only if, $\dim N[\mathfrak{L}(\lambda_0)] = 1$ and λ_0 is a 1-transversal eigenvalue of $\mathfrak{L}$, i.e., if

$$\mathfrak{L}_1 \varphi_0 \notin R[\mathfrak{L}_0], \quad \text{where } \ N[\mathfrak{L}(\lambda_0)] = \mathrm{span}\,[\varphi_0].$$

Consequently, in this particular situation, $\chi[\mathfrak{L};\lambda_0]=1$.

The next result is pivotal in the theory of algebraic multiplicities. It goes back to Chapters 4 and 5 of [13], where the original theory of J. Esquinas and J. López-Gómez [7] and J. Esquinas [6] was considerably polished and sharpened. It was stated as part of Theorem 5.3.1 of [18] by J. López-Gómez and C. Mora-Corral.

Theorem 4.2 *Suppose J is an open subinterval of $\mathbb{R}$, $\mathfrak{L}\in\mathcal{C}^r(J,\mathrm{Fred}_0(U,V))$ for some integer $r\geq 1$, and $\lambda_0\in J$. Then, the following conditions are equivalent:*

(a) *λ_0 is an algebraic eigenvalue of order $1\leq k\leq r$.*

(b) *There exists $\Phi\in\mathcal{C}^\omega(J;\mathrm{Fred}_0(U))$ with $\Phi(\lambda_0)=I_U$ such that λ_0 is a k-transversal eigenvalue of*
$$\mathfrak{L}^\Phi(\lambda):=\mathfrak{L}(\lambda)\Phi(\lambda),\qquad \lambda\in J.$$
Moreover, $\chi[\mathfrak{L}^\Phi;\lambda_0]$ is independent of the transversalizing family of isomorphisms, $\Phi(\lambda)$. Therefore, the next concept of algebraic multiplicity
$$\chi[\mathfrak{L};\lambda_0]:=\chi[\mathfrak{L}^\Phi;\lambda_0]$$
is consistent.

(c) *There exist k finite rank projections $P_j\in\mathcal{L}(U)\setminus\{0\}$, $1\leq j\leq k$, and a map $\mathfrak{M}\in\mathcal{C}^{r-k}(J,\mathrm{Fred}_0(U,V))$, with $\mathfrak{M}(\lambda_0)\in\mathrm{Iso}\,(U,V)$, such that*
$$\mathfrak{L}(\lambda)=\mathfrak{M}(\lambda)[(\lambda-\lambda_0)P_1+I_U-P_1]\cdots[(\lambda-\lambda_0)P_k+I_U-P_k],\quad \lambda\in J. \tag{4.4}$$
Moreover,
$$\chi[\mathfrak{L};\lambda_0]=\sum_{j=1}^k\mathrm{rank}\,P_j, \tag{4.5}$$
independently of the projections P_j chosen.

Based on Theorem 4.2, the next result established the fact that the generalized algebraic multiplicity $\chi[\mathfrak{L};\lambda_0]$ detects any sign jump of $\mathfrak{L}(\lambda)$ at any algebraic eigenvalue λ_0, as discussed by P. Benevieri and M. Furi [3]. It is Theorem 3.3 of J. López-Gómez and C. Mora-Corral [17].

Theorem 4.3 *Suppose J is an open subinterval of $\mathbb{R}$, $\mathfrak{L}\in\mathcal{C}^r(J,\mathrm{Fred}_0(U,V))$ for some integer $r\geq 1$, and $\lambda_0\in J$ is an algebraic eigenvalue of $\mathfrak{L}$ or order $1\leq k\leq r$. Once oriented $\mathfrak{L}$, $\mathrm{sgn}\,\mathfrak{L}(\lambda)$ changes as λ crosses λ_0 if, and only if, $\chi[\mathfrak{L};\lambda_0]$ is odd.*

Therefore, according to Theorem 3.1 of P. Benevieri and M. Furi [3], the next result holds (see Theorem 3.4 of J. López-Gómez and C. Mora-Corral [17]).

Theorem 4.4 *Suppose J is an open subinterval of $\mathbb{R}$, $\mathfrak{L} \in \mathcal{C}^r(J, \mathrm{Fred}_0(U,V))$ for some integer $r \geq 1$, and $\lambda_0 \in J$ is an algebraic eigenvalue of $\mathfrak{L}$ or order $1 \leq k \leq r$ such that $\chi[\mathfrak{L};\lambda_0]$ is odd. Then, $(\lambda_0, 0)$ is a bifurcation point from $(\lambda, 0)$ to a continuum of non-trivial solutions for every continuous function $\mathfrak{F} : \mathbb{R} \times U \to V$ satisfying (F1), (F2) and (F3) with $D_u\mathfrak{F}(\lambda,0) = \mathfrak{L}$.*

The main theorem of J. Esquinas and J. López-Gómez [7] shows that actually Theorem 4.4 is optimal, in the sense that whenever $\chi[\mathfrak{L};\lambda_0]$ is even there exists a smooth $\mathfrak{F}$ satisfying (F1), (F2) and (F3), with $D_u\mathfrak{F}(\lambda,0) = \mathfrak{L}$, for which $(\lambda_0, 0)$ is not a bifurcation point of $\mathfrak{F} = 0$ from $(\lambda, 0)$ (see Chapter 4 of [13]). In particular, the next generalized version of the local bifurcation theorem of M. G. Crandall and P. H. Rabinowitz [5] holds. It should be noted that we are not requiring that $\mathfrak{F}$ is of class $\mathcal{C}^2$. Therefore, in the next result the bifurcating continuum is far from being a $\mathcal{C}^1$ curve, as it occurs in the context of the classical theorem of [5].

Corollary 4.1 *Suppose J is an open subinterval of $\mathbb{R}$, $\mathfrak{L} \in \mathcal{C}^1(J, \mathrm{Fred}_0(U,V))$ and $\lambda_0 \in J$ is a simple eigenvalue $\mathfrak{L}$ in the sense that*

$$\mathfrak{L}'(\lambda_0)\varphi_0 \notin R[\mathfrak{L}(\lambda_0)], \quad \textit{where } N[\mathfrak{L}(\lambda_0)] = \mathrm{span}\,[\varphi_0].$$

Then, $\chi[\mathfrak{L};\lambda_0] = 1$ and hence, $(\lambda_0, 0)$ is a bifurcation point from $(\lambda, 0)$ to a continuum of non-trivial solutions of $\mathfrak{F} = 0$ for every continuous function $\mathfrak{F} : \mathbb{R} \times U \to V$ satisfying (F1), (F2), (F3) and such that $D_u\mathfrak{F}(\lambda,0) = \mathfrak{L}$.

When, in addition, $\mathfrak{F}$ is of class $\mathcal{C}^2$, then the bifurcating continuum consists of a $\mathcal{C}^1$ curve, as established by the theorem of M. G. Crandall and P. H. Rabinowitz [5].

5. A sharp global bifurcation theorem for Fredholm operators

This section reviews the main global bifurcation theorem of J. López-Gómez and C. Mora-Corral [17] and extract some important consequences from it. Given two non-empty subsets of $\mathbb{R}$, A and B, it is said that $A < B$ if $a < b$ for all $(a,b) \in A \times B$. A family, $\mathcal{A}$, whose elements are subsets of a topological space, X, is said to be *locally finite* if for every $x \in X$ there exists a neighborhood, Ω, of x such that $\{A \in \mathcal{A} : A \cap \Omega \neq \emptyset\}$ is finite.

Subsequently, we consider

$$\mathfrak{L}(\lambda) = D_u\mathfrak{F}(\lambda, 0), \qquad \lambda \in \mathbb{R},$$

and $\Sigma = \Sigma(\mathfrak{L})$. The following concept is very useful.

Definition 5.1 *Given $r, s \in \mathbb{Z} \cup \{-\infty, \infty\}$, with $r \leq s$, a family, $\{K_j\}_{j=r}^{s}$, of disjoint closed subsets of $\mathbb{R}$ is said to be* admissible *with respect to Σ if*

$$\Sigma = \bigcup_{j=r}^{s} K_j, \qquad K_j < K_{j+1}, \quad j \in \mathbb{Z} \cap [r, s-1], \tag{5.1}$$

and each of the next conditions is satisfied:

(a) *If $r \in \mathbb{Z}$, then either K_r is compact, or $K_r = (-\infty, a]$ for some $a \in \mathbb{R}$.*

(b) *If $s \in \mathbb{Z}$, then either K_s is compact, or $K_s = [b, +\infty)$ for some $b \in \mathbb{R}$.*

(c) *K_j is compact for all $j \in (r, s) \cap \mathbb{Z}$.*

Naturally, the family $\{K_j\}_{j=r}^s$ must be locally finite, and Σ admits many admissible families, because Σ is a closed subset of $\mathbb{R}$ and any bounded closed subset of $\mathbb{R}$ is compact. In many applications, $\mathfrak{L}(\lambda)$ is real analytic in λ and hence, thanks to Theorem 4.1, either $\Sigma = \mathbb{R}$, or Σ is discrete. Therefore, Σ is discrete if $\mathfrak{L}(a) \in \mathrm{Iso}\,(U, V)$ for some $a \in \mathbb{R}$. In such case, each of the K_j's can be taken as a single point of the spectrum Σ, which is the most common situation covered by the available literature.

Associated to any admissible family of disjoint closed subsets with respect to Σ, $\{K_j\}_{j=r}^s$, there is another locally finite family of open subintervals of $\mathbb{R}$, $\{J_i\}_{i=r-1}^s$, defined by

$$J_i := (\max K_i, \min K_{i+1}), \qquad i \in [r, s-1] \cap \mathbb{Z}, \tag{5.2}$$

if $r = -\infty$ and $s = +\infty$. When $r \in \mathbb{Z}$ and K_r is compact, we should add $J_{r-1} := (-\infty, \min K_r)$ to the previous family. Similarly, when $s \in \mathbb{Z}$ and K_s is compact, $J_s := (\max K_s, +\infty)$ should be added to the previous ones. By construction,

$$J_i \cap \Sigma = \emptyset \qquad \text{for all} \quad i \in \mathbb{Z} \cap [r-1, s]$$

and

$$J_{i-1} < J_i \qquad \text{for all} \quad i \in \mathbb{Z} \cap [r, s].$$

Moreover, the map

$$\bigcup_{i=r-1}^{s} J_i \ni \lambda \mapsto \operatorname{sgn} \mathfrak{L}(\lambda) \in \{-1, 1\}$$

is continuous. Hence, for every $i \in \mathbb{Z} \cap [r-1, s]$, there exists $a_i \in \{-1, 1\}$ such that

$$\operatorname{sgn} \mathfrak{L}(\lambda) = a_i \qquad \text{for all} \ \ \lambda \in J_i.$$

Consequently, a *parity map*, $\mathcal{P}$, associated to the family $\{J_i\}_{i=r-1}^s$, or, equivalently, $\{K_j\}_{j=r}^s$, can defined through

$$\mathcal{P} \ : \ \mathbb{Z} \cap [r-1, s] \to \{-1, 0, 1\}, \qquad \mathcal{P}(i) := \frac{a_i - a_{i-1}}{2}. \tag{5.3}$$

It should be noted that, setting

$$\Gamma_0 := \{i \in \mathbb{Z} \cap [r, s] \ : \ a_{i-1} = a_i\}, \quad \Gamma_1 := \{i \in \mathbb{Z} \cap [r, s] \ : \ a_{i-1} \neq a_i\},$$

the parity $\mathcal{P}$ satisfies the following properties:

- $\mathcal{P}(i) = 0$ if $i \in \Gamma_0$.
- $\mathcal{P}(i) = \pm 1$ if $i \in \Gamma_1$.
- $\mathcal{P}(i)\mathcal{P}(j) = -1$ if $i, j \in \Gamma_1$ with $i < j$ and $(i, j) \cap \Gamma_1 = \emptyset$.

Moreover, any map defined in $\mathbb{Z} \cap [r, s]$ satisfying these properties must be either $\mathcal{P}$ or $-\mathcal{P}$. Thus, either Γ_0, or Γ_1, determines $\mathcal{P}$ up to a change of sign.

Subsequently, we consider a continuous map $\mathfrak{F} : \mathbb{R} \times U \to V$ satisfying (F1), (F2) and (F3), with $\mathfrak{L} = D_u\mathfrak{F}(\cdot, 0)$, and an admissible family with respect to Σ, $\{K_j\}_{j=r}^{s}$, with associated family of open intervals $\{J_i\}_{i=r-1}^{s}$, and we set

$$\mathfrak{S} := \mathcal{S} \cup \bigcup_{j=r}^{s} [K_j \times \{0\}], \tag{5.4}$$

where $\mathcal{S}$ is given by (1.4). The set $\mathfrak{S}$ is refereed to as the set of *non-trivial solutions* of $\mathfrak{F} = 0$ with respect to $\{K_j\}_{j=r}^{s}$. By Lemma 2.1, it consists of the pairs $(\lambda, u) \in \mathfrak{F}^{-1}(0)$ with $u \neq 0$ plus all possible bifurcation points from $(\lambda, 0)$, $\Sigma \times \{0\}$. Since $\mathcal{S}$ and Σ are closed, $\mathfrak{S}$ is closed.

The next result is a straightforward consequence of Theorem 5.4 of J. López-Gómez and C. Mora-Corral [17], whose proof is based on the degree of P. Benevieri and M. Furi [1]-[3] sketched in Section 3. It extends the findings of [15] and [16].

Theorem 5.1 *Suppose $\mathfrak{C}$ is a compact component of $\mathfrak{S}$. Then, the set*

$$\mathcal{B} := \{j \in \mathbb{Z} \cap [r, s] \ : \ \mathfrak{C} \cap (K_j \times \{0\}) \neq \emptyset\}$$

is finite, possibly empty. Moreover,

$$\sum_{i \in \mathcal{B}} \mathcal{P}(i) = 0 \qquad \text{if} \quad \mathcal{B} \neq \emptyset. \tag{5.5}$$

When $\mathcal{B} = \emptyset$, it is said that $\mathfrak{C}$ is an *isola* with respect to the trivial solution $(\lambda, 0)$. The existence of isolas is well documented in the theory of nonlinear differential equations (see, e.g., J. López-Gómez [13, Section 2.5.2], S. Cano-Casanova et al. [4] and J. López-Gómez and M. Molina-Meyer [14]).

When $\mathcal{B} \neq \emptyset$, it is said that $\mathfrak{C}$ *bifurcates from the trivial solution* $(\lambda, 0)$. In this case the set of indices $\mathcal{B}$ provides us with the set of compact subsets, K_j's, of Σ where $\mathfrak{C}$ bifurcates from the trivial state $(\lambda, 0)$. Note that if $r \in \mathbb{Z}$ and $K_r = (-\infty, a]$ for some $a \in \mathbb{R}$, then $r \notin \mathcal{B}$. Indeed, if

$$\mathfrak{C} \cap (K_r \times \{0\}) \neq \emptyset,$$

then $(-\infty, a] \times \{0\} \subset \mathfrak{C}$, because $\mathfrak{C}$ is a closed and connected subset of $\mathbb{R} \times U$ maximal for the inclusion. But this is impossible because $\mathfrak{C}$ is bounded. Therefore, K_j is compact for all $j \in \mathcal{B}$. In particular, this implies that $\mathcal{B}$ is finite.

J. López-Gómez [13, Section 2.5.2] and J. López-Gómez and M. Molina-Meyer [14]) gave a number of examples of compact components, $\mathfrak{C}$, with $\mathcal{B} \neq \emptyset$.

Remark 5.1 *As an immediate consequence from (5.5), when $\mathcal{P}(i) = \pm 1$ for some $i \in \mathcal{B}$, then there exists another $j \in \mathcal{B} \setminus \{i\}$ with $\mathcal{P}(j) = \mp 1$. Therefore, in such case, the component $\mathfrak{C}$ must connect $K_i \times \{0\}$ with $K_j \times \{0\}$. Actually, there is an even number of $i \in \mathcal{B}$'s with $\mathcal{P}(i) = \pm 1$.*

Theorem 5.1 is a substantial generalization of Theorem 6.3.1 of J. López-Gómez [13]. Consequently, it extends to the general setting of Fredholm operators covered by this paper the most pioneering global results of P. H. Rabinowitz [22], L. Nirenberg [20], J. Ize [10] and R. J. Magnus [19] found in the special case when $U = V$ for general classic families of the form

$$\mathfrak{L}(\lambda) = I_U - \lambda T, \qquad T \in \mathcal{K}(U), \tag{5.6}$$

in the context of the local theorem of M. A. Krasnoselskij [12]. Indeed, in Theorem 3.4.1 of of L. Nirenberg [20], attributed to P. H. Rabinowitz there in, L. Nirenberg proved that if $\mathfrak{C}$ is compact, then *" $\mathfrak{C}$ contains a finite number of points $(\lambda_j, 0)$ with $1/\lambda_j$ eigenvalues of T. Furthermore the number of such points having odd multiplicity is even."*

When (5.6) holds, since $\mathfrak{L}(0) = I_U \in \mathrm{Iso}\,(U)$, owing to Theorem 4.1, $\Sigma(\mathfrak{L})$ is discrete and every $\lambda_j \in \Sigma$ must be an algebraic eigenvalue of $\mathfrak{L}$. Moreover, according to Theorem 5.4.1 of J. López-Gómez [13],

$$\chi[\mathfrak{L}; \lambda_j] = m_a(\lambda_j^{-1}; T) := \dim \bigcup_{k=1}^{\infty} N[(\lambda_j^{-1} - T)^k].$$

In other words, $\chi[\mathfrak{L}; \lambda_j]$ equals the classical concept of algebraic multiplicity.

More generally, according to Theorems 4.1 and 4.3, when $\mathfrak{L}(\lambda)$ is a real analytic family of Fredholm operators of index zero such that $\mathfrak{L}(a)$ is an isomorphism for some $a \in \mathbb{R}$, $\Sigma(\mathfrak{L})$ is discrete and if

$$\Sigma = \{\lambda_j \ : \ j \in I\}$$

for some $I \subset \mathbb{Z}$, and we take $K_j = \{\lambda_j\}$ for all $j \in I$, then $\mathcal{P}(j) = \pm 1$ if and only $\chi[\mathfrak{L}; \lambda_j]$ is odd for all $j \in I$. Therefore, according to Theorem 5.1, if

$$\mathfrak{C} \cap (\mathbb{R} \times \{0\}) = \{(\lambda_{i_1}, 0), ..., (\lambda_{i_N}, 0)\},$$

then

$$\sum_{j=1}^{N} \mathcal{P}(\lambda_{i_j}) = 0$$

and consequently, the number of eigenvalues, λ_{i_j}, with an odd multiplicity must be even, much like in the classical case covered by P. H. Rabinowitz [22] and L. Nirenberg [20], though in the general context of this paper, Σ might not be a discrete set and $\mathfrak{F}(\lambda, \cdot)$ is far from being a compact perturbation of the identity map, but a general Fredholm operator of index zero.

6. Some simple consequences from Theorem 5.1

As an immediate consequence, the next generalized version of the *global alternative* of P. H. Rabinowitz [22] holds.

Theorem 6.1 *Suppose $\mathfrak{C}$ is a bounded component of $\mathfrak{S}$ such that*

$$\mathfrak{S} \cap (K_{j_0} \times \{0\}) \neq \emptyset$$

for some $j_0 \in \mathcal{B}$ with $\mathcal{P}(j_0) = \pm 1$. Then, either

(A1) $\mathfrak{C}$ is not compact; or

(A2) there exists another $\mathcal{B} \in j_1 \neq j_0$ with $\mathcal{P}(j_1) = \mp 1$ such that

$$\mathfrak{S} \cap (K_{j_1} \times \{0\}) \neq \emptyset.$$

Consequently, $\mathfrak{S}$ connects $K_{j_0} \times \{0\}$ to $K_{j_1} \times \{0\}$.

As the degree of P. Benevieri and M. Furi extends the concept of parity introduced by P. M. Fitzpatrick and J. Pejsachowicz, Theorem 6.1 of J. Pejsachowicz and P. J. Rabier [21] holds from the previous result.

As another straightforward consequence from Theorem 5.1, the following global version of the local theorem of M. G. Crandall and P H. Rabinowitz [5] holds.

Theorem 6.2 *Suppose $\mathfrak{L} \in \mathcal{C}^1(\mathbb{R}, \mathrm{Fred}_0(U,V))$ and $\lambda_0 \in \mathbb{R}$ is a simple eigenvalue $\mathfrak{L}$, as discussed by M. G. Crandall and P. H. Rabinowitz [5], i.e.,*

$$\mathfrak{L}'(\lambda_0)\varphi_0 \notin R[\mathfrak{L}(\lambda_0)], \quad \textit{where} \;\; N[\mathfrak{L}(\lambda_0)] = \mathrm{span}\,[\varphi_0].$$

Then, for every continuous function $\mathfrak{F} : \mathbb{R} \times U \to V$ satisfying (F1), (F2), (F3) and such that $D_u\mathfrak{F}(\cdot,0) = \mathfrak{L}$, $(\lambda_0,0)$ is a bifurcation point from $(\lambda,0)$ to a continuum of non-trivial solutions of $\mathfrak{F} = 0$.

For any of these $\mathfrak{F}$'s, let $\{K_j\}_{j=r}^{s}$ be an admissible family for Σ with $K_0 = \{\lambda_0\}$, and let $\mathfrak{C}$ be the component of $\mathfrak{S}$ such that $(\lambda_0,0) \in \mathfrak{C}$. Then, some of the following alternatives holds:

(a) *$\mathfrak{C}$ is not compact.*

(b) *There exists $\Sigma \ni \lambda_1 \neq \lambda_0$ such that $(\lambda_1,0) \in \mathfrak{C}$.*

Actually, if $\mathfrak{C}$ is compact, then there exists $N \geq 1$ such that

$$(K_j \times \{0\}) \cap \mathfrak{C} \neq \emptyset \quad \textit{if, and only if,} \quad j \in \{j_{i_1},...,j_{i_N}\} \subset \mathbb{Z} \cap [r,s]$$

with $j_{i_k} = 0$ for some $k \in \{1,...,N\}$. Moreover,

$$\sum_{k=1}^{N} \mathcal{P}(j_{i_k}) = 0.$$

Therefore, $\mathfrak{C}$ connects $(\lambda_0, 0)$ with an odd number of $K_j \times \{0\}$'s with parity ± 1, where the sign of the orientation changes.

Proof. According to Definition 4.2, λ_0 is a 1-transversal eigenvalue of $\mathfrak{L}(\lambda)$ with $\chi[\mathfrak{L}; \lambda_0] = 1$. Thus, by Theorem 4.2, λ_0 must be an algebraic eigenvalue of $\mathfrak{L}(\lambda)$ of order one, as discussed by Definition 4.1. In particular, $\mathfrak{L}(\lambda) \in \mathrm{Iso}\,(U, V)$ for λ sufficiently close to λ_0 with $\lambda \neq \lambda_0$. Thus, since, $\chi[\mathfrak{L}; \lambda_0] = 1$, $\operatorname{sgn} \mathfrak{L}(\lambda)$ changes of sign as λ crosses λ_0 and hence, if $\{K_j\}_{j=r}^{s}$ is an admissible family for Σ with $K_0 = \{\lambda_0\}$, we have that $\mathcal{P}(0) = \pm 1$. The remaining assertions of the theorem are immediate consequences of Theorem 6.1. □

References

[1] P. Benevieri, M. Furi: *A simple notion of orientability for Fredholm maps of index zero between Banach manifolds and degree theory.* Ann. Sci. Math. Québec **22** (1998) 131–148.

[2] P. Benevieri, M. Furi: *On the concept of orientability for Fredholm maps between real Banach manifolds.* Top. Methods in Nonl. Anal. **16** (2000) 279–306.

[3] P. Benevieri, M. Furi: *Bifurcation results of families of Fredholm maps os index zero between Banach spaces.* In *Nonlinear Analysis and its Applications (St. John's, NF, 1999),* Nonlinear Anal. Forum **6** (2001) 35–47.

[4] S. Cano-Casanova, J. López-Gómez, M. Molina-Meyer: *Isolas: compact solution components separated away from a given equilibrium state.* Hiroshima J. of Mathematics **34** (2004) 177–199.

[5] M. G. Crandall, P. H. Rabinowitz: *Bifurcation from simple eigenvalues.* J. Functional Analysis **8** (1971) 321–340.

[6] J. Esquinas: *Optimal multiplicity in local bifurcation theory, II: General case.* J. Differential Equations **75** (1988) 206–215.

[7] J. Esquinas, J. López-Gómez: *Optimal multiplicity in local bifurcation theory, I: Generalized generic eigenvalues.* J. Differential Equations **71** (1988) 72-92.

[8] P. M. Fitzpatrick, J. Pejsachowicz: *Parity and generalized multiplicity.* Trans. Amer. Math. Soc. 326 (1991) 281–305.

[9] P. M. Fitzpatrick, J. Pejsachowicz: *Orientation and the Leray-Schauder theory for fully nonlinear elliptic boundary value problems.* Memoirs Amer. Math. Soc. 483, 1993.

[10] J. Ize: *Bifurcation Theory for Fredholm operators.* Memoirs Amer. Math. Soc. 174, Prvidence, RI, 1976.

[11] T. Kato: *Perturbation Theory for Linear Operators.* Grundlehren Math. Wiss. 132, Springer, Berlin, 1980.

[12] M. A. Krasnoselskij: *Topological Theory in the Theory of Nonlinear Integral Equations.*

[13] J. López-Gómez: *Spectral Theory and Nonlinear Functional Analysis.* Research Notes in Mathematics 426, Chapman and Hall/CRC Press, Boca Raton, 2001.

[14] J. López-Gómez, M. Molina-Meyer: *Bounded components of positive solutions of abstract fixed point equations: mushrooms, loops and isolas.* Journal of Differential Equations **209** (2005) 416–441.

[15] J. López-Gómez, C. Mora-Corral: *Minimal complexity of semi-bounded components in bifurcation theory.* Nonlinear Analysis **58** (2004) 749–777.

[16] J. López-Gómez, C. Mora-Corral: *Counting solutions of nonlinear abstract equations.* Topological Methods in Nonlinear Analysis **24** (2004) 307–335.

[17] J. López-Gómez, C. Mora-Corral: *Counting zeros of C^1 Fredholm maps of index 1.* Bull. London Math. Soc. **37** (2005) 778–792.

[18] J. López-Gómez, C. Mora-Corral: *Algebraic Multiplicity of Eigenvalues of Linear Operators.* Operator Theory, Advances and Applications Vol. 177, Birkhäuser, 2007.

[19] R. J. Magnus: *A generalization of multiplicity and the problem of bifurcation.* Proc. London Mathematical Society **32** (1976) 251–278.

[20] L. Nirenberg: *Topics in Nonlinear Functional Analysis.* Lectures Notes of the Courant Institute of Mathematical Sciences, NYU, 1974.

[21] J. Pejsachowicz, P. J. Rabier: *Degree theory for $\mathcal{C}^1$ Fredholm mappings of index zero.* J. Annal. Math., **76** (1998) 289–319.

[22] P. H. Rabinowitz: *Some global results for nonlinear eigenvalue problems.* J. of Functional Analysis **7** (1971) 487–513.

Pascal's triangle, Stirling numbers and the Euler characteristic

Ana Luzón[*], Manuel A. Morón[*] and Felipe Prieto-martínez[*]

Departamento de Matemática Aplicada
Universidad Politécnica de Madrid
28040 Madrid, Spain
anamaria.luzon@upm.es

Departamento de Geometría y Topología
Universidad Complutense de Madrid
28040 Madrid, Spain
mamoron@mat.ucm.es

Departamento de Matemáticas
Universidad Autónoma de Madrid
28049 Madrid, Spain
luisfelipe.prieto@uam.es

This note is dedicated to Jose María Montesinos: we will always be indebted to you.

ABSTRACT

In this short note we use some basic properties of the classical families of binomial and Stirling numbers and some linear algebraic tools to get results relating those families to the Euler characteristic of simplicial complexes. In fact, we identify invariance in non-topological settings establishing later uniqueness results in the topological one.

2010 Mathematics Subject Classification: 11Z05, 05E45, 15A18.
Key words: Pascal triangle, Stirling numbers, Euler characteristic.

[*]Authors partially supported by project MINECO MTM2012-30719

1. Introduction

According to D. Eppstein in [4] and the introduction of N. Levitt in [8], the Euler characteristic χ is the best known as well as the most ancient topological invariant. Moreover, the Euler formula $E - V + F = 2$ is one of many theorems in mathematics which are important enough as to be proved repeatedly in surprisingly many different ways.

On this line the unique invariance of the Euler characteristic, among linear combinations on the numbers of faces of triangulations, is known and reproved from time to time in different realms. Up to our knowledge, the first proof appeared in Mayer [11]. More recently in [8], [5] and [12] there are some related results. Very recently in [16], this result is strengthened, in the framework of combinatorial manifolds, to non-linear functions on the number of faces.

The simple observation that the f-vectors of n-simplices, considered as abstract simplicial complexes, can be placed forming an infinite lower triangular matrix which is almost the Pascal triangle, allowed us to use the ideas in our previous works about Riordan matrices, see for example [9] and [10]. These kind of matrices and the group structure were introduced in [13] and [14].

Our main idea is to consider the $\mathbb{K}$-linear action induced by such a matrix of f-vectors on $\mathbb{K}[[x]]$, where we consider the natural $\mathbb{K}$-linear space structure on the set of formal power series with coefficients on a field $\mathbb{K}$ of characteristic zero. Using only this matrix we obtain, in particular, the unique homotopy invariance of the Euler characteristic among all possible linear combinations on components of f-vectors.

After that, we use the known formula of the change of f-vectors under barycentric subdivisions, involving Stirling numbers, to prove that the multiples of the Euler characteristic are the unique invariant under barycentric subdivisions in the class of all n-simplices. Consequently, we prove the unique topological invariance of the Euler characteristic among all possible linear combinations on components of f-vectors.

We think that the main original result in this note is Theorem 8, where we give a description of the eigenspace associated to the eigenvalue 1 in the linear action induced by an infinite matrix describing the variation of f-vectors under barycentric subdivisions. This is the key tool to get invariance in the non-topological setting and to induce unique invariance in the topological one.

Probably almost all results herein are known. Certainly those in the topological framework are, but we think that the pointed out connections between the three classical objects in the title can still be of interest.

2. Some new definitions

We suppose that the basic definitions of abstract and geometric finite simplicial complexes are known. See the corresponding introductory chapters in [15] or [17]. We call herein the f-vector of an n-dimensional finite simplicial complex $\mathcal{F}$ to $(f_0, f_1, \ldots, f_n)$, where f_i counts the number of i-faces of $\mathcal{F}$. So f_0 is the number of vertices, f_1 is the number of edges, and so on. We have to note that in other places the f-vector includes $f_{-1} = 1$ as its first coordinate which corresponds to the interpretation of the empty set as the unique (-1)-dimensional face. Given a m-dimensional simplicial complex $\mathcal{F}$ with f-vector $f^{\mathcal{F}} = \left(f_0^{\mathcal{F}}, f_1^{\mathcal{F}}, \ldots, f_m^{\mathcal{F}}\right)$ the Euler characteristic is defined by

$$\chi(\mathcal{F}) = \sum_{k=0}^{m} (-1)^k f_k^{\mathcal{F}}.$$

Consider the geometric realization $|\mathcal{F}|$ of $\mathcal{F}$. It is known that $\chi(\mathcal{F})$ depends on $|\mathcal{F}|$ but not on the triangulation $\mathcal{F}$. Even more, $\chi(\mathcal{F})$ depends only on the homotopy type of $|\mathcal{F}|$ because it can be expressed in terms of the ranks of the homology groups $H_k(|\mathcal{F}|)$ which are homotopy invariants. This previous result is the so called Euler-Poincaré formula. See page 146 in [7].

Note that $\chi(\mathcal{F})$ can be expressed as the following product of infinite matrices :

$$\chi(\mathcal{F}) = (f_0^{\mathcal{F}}, f_1^{\mathcal{F}}, \ldots, f_m^{\mathcal{F}}, 0, \ldots) \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \\ (-1)^m \\ \vdots \end{pmatrix}$$

and the column matrix does not depend on the complex $\mathcal{F}$. The generating function of this column matrix is $\dfrac{1}{1+x}$

A natural way to define linear combinations on the number of faces indepently on the polyhedron even on its dimension is the following

Definition 1 *Let $\gamma(x) = \sum_{n \geq 0} \gamma_n x^n$ be any power series with coefficients in $\mathbb{K}$. Suppose that $\mathcal{F}$ is a finite simplicial complex with f-vector $(f_0^{\mathcal{F}}, \ldots, f_m^{\mathcal{F}}, 0, \ldots)$, we define the linear combination induced by the series γ, and denote it by $\chi(\gamma, \mathcal{F})$, as*

$$\chi(\gamma, \mathcal{F}) = \sum_{k=0}^{m} \gamma_k f_k^{\mathcal{F}}.$$

So, the Euler characteristic is

$$\chi\left(\frac{1}{1+x}, \mathcal{F}\right).$$

The Euler characteristic is, on one hand, a homotopy invariant in the class of finite polyhedra and, on the other hand, it is a linear combination on the number of faces of any triangulation of the polyhedron.

Let us define

Definition 2 *Let $\gamma \in \mathbb{K}[[x]]$, $\gamma(x) = \sum_{n\geq 0} \gamma_n x^n$ be any power series. We say that*

(a) $\gamma(x)$ is a homotopy invariant for the class of finite simplicial complexes if given any two of them $\mathcal{F}_1$ and $\mathcal{F}_2$, then $\chi(\gamma, \mathcal{F}_1) = \chi(\gamma, \mathcal{F}_2)$ provided $|\mathcal{F}_1|$ and $|\mathcal{F}_2|$ have the same homotopy type.

(b) $\gamma(x)$ is a topological invariant for the class of finite simplicial complexes if given any two of them $\mathcal{F}_1$ and $\mathcal{F}_2$, then $\chi(\gamma, \mathcal{F}_1) = \chi(\gamma, \mathcal{F}_2)$ provided $|\mathcal{F}_1|$ and $|\mathcal{F}_2|$ are homeomorphic.

Of course, any series which is a homotopy invariant is a topological invariant. We can restrict the above definition to any subclass of finite simplicial complexes.

3. Pascal's Triangle and the Euler characteristic

The basic pieces to construct polyhedra in Topology are the geometric n-simplices, $|\Delta_n|$. Topologically they can be described as the convex hull of $n+1$ affinely independent points in a suitable euclidean space.

The abstract description of an n-simplex as a simplicial complex, denoted by Δ_n, is given by considering all the non-empty subsets of a set of vertices $V = \{v_0, \ldots, v_n\}$ with $n+1$ points. So, the corresponding f-vectors are easily computed using combinatorial numbers. If we denote by f^{Δ_n} the f-vector of Δ_n we get

$$f^{\Delta_n} = \left(\binom{n+1}{1}, \binom{n+1}{2}, \ldots, \binom{n+1}{n+1}\right).$$

We can place these vectors f^{Δ_n} forming an infinite lower triangular matrix,

$$F = \left(\binom{n+1}{k+1}\right)_{n,k\geq 0}.$$

We note that the non-null part coincides with Pascal's triangle without the first row and column.

This matrix as well as the corresponding matrix representation of Pascal's triangle, and some of its generalizations, are elements of a group under the usual product of matrices. This group is known as the Riordan group. We approached this group in [10], see also [9], using the Banach Fixed Point Theorem. To describe the elements $T(\beta \mid \alpha)$ in this group as in [10], we use a pair of formal power series $\alpha(x) = \sum_{n\geq 0} \alpha_n x^n$ and $\beta(x) = \sum_{n\geq 0} \beta_n x^n$ such that $\alpha_0 \neq 0$ and $\beta_0 \neq 0$. The columns of $T(\beta \mid \alpha)$ are the coefficients of the elements in the geometric progression, in $\mathbb{K}[[x]]$, whose first term is the series $\dfrac{\beta(x)}{\alpha(x)}$ and the ratio is the series $\dfrac{x}{\alpha(x)}$.

The representations of the product and the inverse in this group are:

$$T(\beta \mid \alpha)T(\bar{\beta} \mid \bar{\alpha}) = T(\tilde{\beta} \mid \tilde{\alpha}),$$

where

$$\tilde{\beta}(x) = \beta(x)\bar{\beta}\left(\frac{x}{\alpha(x)}\right), \qquad \tilde{\alpha}(x) = \alpha(x)\bar{\alpha}\left(\frac{x}{\alpha(x)}\right)$$

$$(T(\beta \mid \alpha))^{-1} = T\left(\frac{1}{\beta(\omega^{-1})}\middle|\frac{1}{\alpha(\omega^{-1})}\right), \qquad \omega = \frac{x}{\alpha}, \qquad \omega \circ \omega^{-1} = \omega^{-1} \circ \omega = x.$$

Besides, we can consider the matrix $T(\beta \mid \alpha)$, like in Linear Algebra, as the associated matrix to a $\mathbb{K}$-linear isometry for a suitable ultrametric d, see [10], defined by:

$$\begin{array}{rccc} T(\beta \mid \alpha): & (\mathbb{K}[[x]], d) & \to & (\mathbb{K}[[x]], d) \\ & \gamma & \mapsto & T(\beta \mid \alpha)(\gamma) = \frac{\beta}{\alpha}\gamma\left(\frac{x}{\alpha}\right) \end{array} \tag{3.1}$$

In these terms Pascal's triangle is $T(1 \mid 1 - x)$ and our matrix of f-vectors is $F = T\left(\frac{1}{1-x}\middle|1-x\right)$, because we obtain F from Pascal's triangle by deleting the first row and column, see page 3614 in [9].

Using, essentially, Pascal's triangle we get

Proposition 3 *The unique linear combinations of the number of faces which assign the same number to every abstract n-simplex are the multiples of the Euler characteristic.*

Proof. To calculate the Euler characteristic $\chi(\Delta_n)$ $\forall n \geq 0$, we only need to compute

$$T\left(\frac{1}{1-x}\middle|1-x\right)\left(\frac{1}{1+x}\right) = \sum_{n\geq 0} \chi(\Delta_n)x^n,$$

because the rows of the matrix F are the vectors f^{Δ_n}.

On the other hand, using (3.1) we get

$$T\left(\frac{1}{1-x}\middle|1-x\right)\left(\frac{1}{1+x}\right) = \frac{1}{(1-x)^2}\frac{1}{1+\frac{x}{1-x}} = \frac{1}{1-x},$$

then

$$\sum_{n\geq 0} \chi(\Delta_n)x^n = \frac{1}{1-x} \qquad \text{equivalently} \qquad \chi(\Delta_n) = 1 \quad \forall n \geq 0,$$

as everybody knows. Finally the linearity of the action of $F = T\left(\frac{1}{1-x}\middle|1-x\right)$ on $\mathbb{K}[[x]]$ implies the result for the multiples.

Suppose now a series $\gamma(x) = \sum_{n\geq 0} \gamma_n x^n$ and $k \in \mathbb{K}$ such that $\chi(\gamma, \Delta_n) = k \ \forall n \geq 0$. Consider the Riordan matrix $T\left(\frac{1}{1-x}\middle|1-x\right)$ whose rows are the f-polynomials f^{Δ_n}. Then

$$T\left(\frac{1}{1-x}\middle|1-x\right)(\gamma(x)) = \frac{k}{1-x},$$

Now, in the Riordan group,

$$T^{-1}\left(\frac{1}{1-x}\middle|1-x\right) = T\left(\frac{1}{1+x}\middle|1+x\right),$$

consequently

$$\gamma(x) = T\left(\frac{1}{1+x}\middle|1+x\right)\left(\frac{k}{1-x}\right) = \frac{k}{1+x}$$

and the proof is finished. □

Remark 4 *The fact that the Euler characteristics of $\chi(\Delta_n)$ are equal $\forall n \geq 0$ is not a topology matter in the sense that we do not need any topological argument to prove that. In fact it can be considered as a consequence of the Newton binomial formula.*

We also have the following topological consequence

Corollary 5 *The multiples of the Euler characteristic are the unique linear combinations on the number of faces of triangulations, of finite simplicial complexes, which are homotopy invariant.*

Proof. Consider the geometric realizations $|\Delta_n|$ of Δ_n. All of them are homotopically equivalent because they are contractible. If we have a linear combination on the number of faces of f-vectors of triangulations which is homotopy invariant, it should assign the same number to all the triangulations Δ_n and then it is a multiple of the Euler characteristic. □

Remark 6 *(a) The same proof is valid in the more restrictive framework of simple homotopy theory, see [2].*

4. Stirling numbers and the unique invariance of the Euler characteristic.

In [1] the authors treat f-vectors of barycentric subdivision of simplicial complexes to get, in particular, that certain limiting behavior depending on the iteration of the barycentric subdivision of a simplicial complex, does not depend on the complex itself but on the dimension of such complex, see also [3]. In that paper a formula for the variation of the f-vector after a barycentric subdivision is given.

Now we reproduce the formula at page 850 in [1] taking into account that we use the f-vector $(f_0, f_1, \ldots, f_m)$ and not the extended f-vector $(f_{-1}, f_0, f_1, \ldots, f_m)$.

Proposition 7 *Let $\mathcal{F}$ be a m-dimensional simplicial complex. Denote by $sd(\mathcal{F})$ the complex obtained by the barycentric subdivision of $\mathcal{F}$. Then*

$$f_j^{sd(\mathcal{F})} = \sum_{i=0}^{m} f_i^{\mathcal{F}} (j+1)! \left\{ {i+1 \atop j+1} \right\} \qquad \textit{for} \qquad j = 0, \ldots, m.$$

In the above result $\left\{ {k \atop l} \right\}$ represents the corresponding Stirling number of the second kind as denoted in [6] Chapter 6.

Let $B = (b_{i,j})_{i,j \geq 0}$ be the matrix with $b_{i,j} = (j+1)! \left\{ {i+1 \atop j+1} \right\}$ $i, j \geq 0$. The formula in the proposition above converts to

$$\begin{aligned}
&(f_0^{sd(\mathcal{F})}, f_1^{sd(\mathcal{F})}, \ldots, f_m^{sd(\mathcal{F})}, 0, \ldots) \\
&= (f_0^{\mathcal{F}}, f_1^{\mathcal{F}}, \ldots, f_m^{\mathcal{F}}, 0, \ldots) \begin{pmatrix} \left\{ {1 \atop 1} \right\} & & & & \\ \left\{ {2 \atop 1} \right\} & 2! \left\{ {2 \atop 2} \right\} & & & \\ \vdots & \vdots & \ddots & & \\ \left\{ {n \atop 1} \right\} & 2! \left\{ {n \atop 2} \right\} & \cdots & n! \left\{ {n \atop n} \right\} & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.
\end{aligned}$$

By the usual way B induces a $\mathbb{K}$-linear isomorphism $B : \mathbb{K}[[x]] \to \mathbb{K}[[x]]$, $B(\zeta(x)) = \eta(x)$, such that if

$$\zeta(x) = \sum_{n \geq 0} \zeta_n x^n, \qquad \text{and} \qquad \eta(x) = \sum_{n \geq 0} \eta_n x^n,$$

then

$$B(\zeta_n) = (\eta_n),$$

where (ζ_n), (η_n) are considered as column vectors. For this operator we get

Theorem 8 *The number 1 is an eigenvalue for the operator B. Moreover the eigenspace associated to 1 is $\left\{ \dfrac{k}{1+x},\ k \in \mathbb{K} \right\}$.*

Proof. If $D = (d_{i,j})$ is the diagonal matrix with $d_{i,i} = (i+1)!$ and $S = \left(\left\{ {i+1 \atop j+1} \right\} \right)_{i,j \in \mathbb{R}^{\mathbb{N}}}$ is the matrix of the Stirling numbers of second kind, then $B = SD$. Consider

$$\delta(x) = D\left(\frac{1}{1+x}\right) = \sum_{i \geq 0} (i+1)!(-x)^i,$$

then

$$B\left(\frac{1}{1+x}\right) = \frac{1}{1+x} \quad \Leftrightarrow \quad S(\delta) = \frac{1}{1+x}.$$

Recall that

$$S^{-1} = \left((-1)^{i-j} \left[{i+1 \atop j+1} \right]\right)_{i,j \in \mathbb{R}^{\mathrm{N}}}$$

where $\left[{i \atop j} \right]$ denote the Stirling numbers of the first kind. Since

$$\left[{i \atop 0} \right] = 0 \qquad \forall i \geq 1 \qquad \text{and} \qquad \sum_{j=0}^{i} \left[{i \atop j} \right] = i!,$$

we have

$$S^{-1}\left(\frac{1}{1+x}\right) = \delta(x)$$

and then

$$B\left(\frac{1}{1+x}\right) = \frac{1}{1+x}.$$

See pages 259-264 in [6] for the properties of Stirling numbers used above.

So we have proved that 1 is an eigenvalue and that $\dfrac{1}{1+x}$ is an associated eigenvector, and then $\left\{\dfrac{k}{1+x},\ k \in \mathbb{K}\right\}$ is contained in the corresponding eigenspace.

We consider the finite central $(m+1) \times (m+1)$ submatrices

$$B_m = (b_{i,j})_{i,j=0\ldots m}$$

of the infinite matrix B. Note that, in every B_m the eigenspace associated to the eigenvalue 1 has always dimension 1, because the entries in the main diagonal are all different. To prove that there is not any other eigenvector for B associated to 1, we suppose that $\eta(x) = \sum_{n \geq 0} \eta_n x^n \neq \dfrac{k}{1+x}$, $k \in \mathbb{K}$ is one such eigenvector. This means that there is an $l \in \mathbb{R}^{\mathrm{N}}$, $l \geq 1$ such that

$$\sum_{j=0}^{l} \eta_j x^j \neq k \sum_{j=0}^{l} (-x)^j \qquad \text{for any} \qquad k \in \mathbb{K}.$$

So, $(\eta_j)_{j=0,\ldots,l}$, $((-1)^j)_{j=0,\ldots,l}$ are eigenvector associated to 1 for the matrix B_l. This is impossible because they are linearly independent. $\square$

The first consequence we obtain is:

Corollary 9 *The invariance of the Euler characteristic by successive barycentric subdivisions is not a topology matter in the sense that we do not need any topological argument to prove that.*

Proof. Let $\mathcal{F}$ be a finite simplicial complex. Denote by $sd^{(k)}(\mathcal{F})$ the k-th barycentric subdivision of $\mathcal{F}$. Using Proposition 7 we get

$$f^{sd^{(k)}(\mathcal{F})} = f^{\mathcal{F}} B^k.$$

So

$$\chi(sd^{(k)}(\mathcal{F})) = \chi\left(\frac{1}{1+x}, sd^{(k)}(\mathcal{F})\right) = \chi\left(B^k\left(\frac{1}{1+x}\right), \mathcal{F}\right) = \chi\left(\frac{1}{1+x}, \mathcal{F}\right) = \chi(\mathcal{F}).$$

□

In the above proof we only used that $\dfrac{1}{1+x}$ is an eigenvector for the matrix B, and then for B^k, associated to the eigenvalue 1.

As a consequence of the fact that $\left\{\dfrac{1}{1+x}\right\}$ is a base for the eigenspace associated to 1 we get another topological consequence.

Corollary 10 *The unique linear combinations which are invariants under barycentric subdivisions in the class of all n-simplices, Δ_n, are the multiples of the Euler characteristic. In particular, $\dfrac{k}{1+x}$ are the unique series which are topologically invariants in the class of finite simplicial complexes.*

Proof. Let $\gamma(x) = \sum_{n\geq 0} \gamma_n x^n$ be a series which is invariant under barycentric subdivisions in the class of all dimensional simplices. This implies that

$$T\left(\frac{1}{1-x}\middle| 1-x\right)(\gamma) = T\left(\frac{1}{1-x}\middle| 1-x\right) B(\gamma),$$

which is equivalent to

$$\gamma = B(\gamma)$$

because of injectivity of the action of $F = T\left(\dfrac{1}{1-x}\middle| 1-x\right)$. Consequently

$$\gamma = \frac{k}{1+x} \quad \text{for some} \quad k \in \mathbb{K}.$$

The second part follows immediately because the polyhedra $|\mathcal{F}|$ and $|sd(\mathcal{F})|$ are always homeomorphic. □

Remark 11 *The above proof can be also applied to the more restrictive framework of PL-Topology.*

References

[1] F. Brenti, V. Welker: *f-Vectors of barycentric subdivisions.* Math. Z. **259** (2008) 849–865.

[2] M. M. Cohen: *A Course in Simple-Homotopy Theory.* GTM **10**, Springer-Verlag 1973.

[3] E. Dellucchi, A. Pixton, L. Sabalka: *Face vectors of subdivided simplicial complexes.* Discrete Math. **312** (2012) 248–257.

[4] D. Eppstein: *Nineteen Proofs of Euler's Formula: V-E+F=2.* http://www.ics.uci.edu/ eppstein/junkyard/euler/

[5] R. Forman: *The Euler Characteristic Is the unique locally determined numerical invariant of finite simplicial complexes which assigns the same sumber to every cone.* Discrete Comput. Geom. **23** (2000) 485–488.

[6] R. Graham, D. Knuth, O. Patashnik: *Concrete Mathematics.* Addison-Wesley, 1989.

[7] A. Hatcher: *Algebraic Topology.* Cambridge University Press, 2002.

[8] N. Levitt: *The Euler Characteristic Is the unique locally determined numerical homotopy invariant of finite complexes.* Discrete Comput. Geom. **7** (1992) 59–67.

[9] A. Luzón: *Iterative processes related to Riordan arrays: The reciprocation and the inversion of power series.* Discrete Math. **310** (2010) 3607–3618.

[10] A. Luzón, M. A. Morón. *Ultrametrics, Banach's fixed point theorem and the Riordan group.* Discrete Appl. Math. **156** (2008) 2620–2635.

[11] W. Mayer: *A new homology theory. II* Ann. of Math. **43(3)** (1942) 594–605.

[12] J. Roberts *Unusual formulae for the Euler characteristic.* J. Knot Theory Ramifications **11** (2002), 793-796.

[13] L. W. Shapiro, S. Getu, W. J. Woan, L. Woodson: *The Riordan group.* Discrete Appl. Math. **34** (1991) 229–239.

[14] R. Sprugnoli: *Riordan arrays and combinatorial sums.* Discrete Math. **132** (1994) 267–290.

[15] R. P. Stanley: *Combinatorics and Commutative Algebra.* Progress in Mathematics. **41** Birkhäuser, Boston 1996.

[16] L. Yu: *A property that characterizes Euler Characteristic among invariants of combinatorial manifolds.* Adv. in Math. **225** (2010) 794–804.

[17] G. M. Ziegler: *Lectures on Polytopes.* GTM **152** Springer-Verlag, New York 1995.

Probabilidades en espacios topológicos

Juan MARGALEF-ROIG*, Salvador MIRET-ARTÉS*
y Enrique OUTERELO-DOMÍNGUEZ

Instituto de Física Fundamental
Consejo Superior de Investigaciones Científicas
Serrano 123. 28006 Madrid, España.
margalef@imaff.cfmac.csic.es s.miret@imaff.cfmac.csic.es

Departamento de Geometría y Topología
Universidad Complutense de Madrid
28040 Madrid, España.
eouterel@ucm.es

Al prominente topólogo-geómetra Dr. D. José María Montesinos Amilibia con motivo de su jubilación como Catedrático de la Universidad Complutense de Madrid.

Resumen

En este artículo presentamos algunos resultados sobre la teoría de probabilidades en espacios topológicos, que generalizan las propiedades establecidas en espacios Euclideanos y tienen importantes aplicaciones, entre otros, en los campos de la Mecánica Estadística, de la Mecánica Cuántica y de la Economía Financiera.

2010 Mathematics Subject Classification: 60A10, 60B05, 60G05.
Key words: Sigma álgebras, medidas de Borel, probabilidades, procesos estocásticos.

1. Introducción

La Teoría de Probabilidades consiste en el estudio de los espacios de probabilidad y de las variables aleatorias, que matemáticamente se definen como aplicaciones medibles de un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ en un espacio medible $(E, \mathcal{E})$. Al desarrollar la teoría y tratar cuestiones prácticas, se revela como muy importante la versatilidad de la estructura definida por la σ-álgebra $\mathcal{E}$ en E, considerando en algunas ocasiones una estructura topológica subyacente.

*Autores financiado por el Proyecto FIS2011-29596-C02-C01 (Ministerio de Economía y Competitividad)

El movimiento Browniano juega un papel destacado en las aplicaciones de la teoría de procesos estocásticos (véase la sección **6**) a la Mecánica Estadística, a la Mecánica Cuántica y a la Economía Financiera. El descubrimiento del movimiento Browniano se atribuye al botánico inglés Robert Brown (1773-1858) en un artículo publicado en 1828. Sin embargo, el propio Brown en ese artículo cita como precursor a A. Van Leeuwenhoek (1632-1723), y en una publicación de 1829 cita hasta 10 precursores más (entre ellos, Buffon, Spallanzani y Bywater). Brown observó el movimiento de partículas de polen en el agua, y se llega a la primera interpretación dinámica del movimiento Browniano: Las partículas, en un medio denso, tienen actividad.
La primera formulación matemática del movimiento Browniano se debe a Louis Bachelier (1870-1946) que en su tesis doctoral *Théorie de la spéculation*, dirigida por H. Poincaré y presentada el 29 de marzo de 1900, introduce lo que ahora se llama movimiento Browniano lineal para describir el precio de las acciones financieras (este trabajo está influido por la obra *Calcul des changes et philosophie de la bourse* de Jules Regnault (1834-1894), que establece en términos literarios el marco conceptual de la aplicación del cálculo de probabilidades a las operaciones bursátiles). Aparte de los resultados de matemática financiera, Bachelier desarrolló un importante estudio del movimiento Browniano y todo ello cinco años antes de la publicación del famoso trabajo de A. Einstein (1879-1955), publicado en Annalen der Physik, 17 (1905), 549-560. El trabajo de Einstein está motivado por los trabajos experimentales de J. Perrin (1870-1942), y en él se describe el movimiento de una partícula a lo largo de una recta.

La formulación matemática definitiva del movimiento Browniano se debe a Norbert Wiener (1894-1964), en la década de los años 20 del siglo pasado. En su teoría Wiener introduce una medida en el espacio de las funciones continuas sobre $[0,1]$ que dejan invariante el origen y la σ-álgebra generada por los cilindros de dicho espacio, y la integral de los funcionales sobre este espacio relativa a esta medida se denomina integral de Wiener. Esta integral fue generalizada por K. Itô (1915-2008) en 1944, introduciendo el cálculo estocástico y el movimiento Browniano geométrico. Este movimiento resultó ser de gran importancia para modelizar la evolución de los precios de las acciones en mercados financieros. Paul A. Samuelson (1915-2009) desarrolló, desde 1965, el estudio del movimiento Browniano geométrico en conexión con la economía y recibió el premio Nobel de Economía en 1970 por estas aportaciones. En 1973, Fischer Black (1938-1995) y Myron S. Scholes (1941-*) e, independientemente Robert C. Merton (1944-*), utilizaron el movimiento Browniano geométrico para asignar precio a las opciones (contrato que da derecho a su poseedor a vender o comprar un activo financiero a un precio determinado durante un periodo de tiempo o en una fecha determinada), creando el llamado, hoy día, modelo BSM (de las iniciales de los tres autores anteriores; Scholes y Merton fueron galardonados con el premio Nobel de Economía en 1997 (Black había fallecido dos años antes)) de los mercados financieros continuos.

Por otra parte, en la Física Estadística, el movimiento Browniano es el paradigma de proceso estocástico. Inicialmente, se parte del modelo sencillo del paseo aleatorio, se deduce una ecuación maestra para llegar a la ecuación de Fokker-Planck, y por la técnica de convergencia en distribución se alcanza el movimiento Browniano. Richard Feynman (1918-1988) en su formulación alternativa de la Mecánica Cuántica introduce la integral de camino (la denominada integral de camino de Feynman) íntimamente relacionada con la integral de Wiener, y estas ideas fueron aplicadas por Mark Kac (1914-1984) a la resolución de un tipo especial de ecuaciones en derivadas parciales parabólicas. Esto es un ejemplo de la influencia recíproca de la Matemática y los problemas de la Física.

En el presente artículo vamos a considerar el caso en que el espacio medible $(E, \mathcal{E})$ proceda de un espacio topológico $(X, \mathcal{T})$, tomando $E = X$ y $\mathcal{E} = \sigma^X(\mathcal{T})$, la σ-álgebra generada por sus subconjuntos abiertos (conjuntos de Borel), e indicaremos cómo la teoría desarrollada permite, entre otras muchas aplicaciones, dar una formalización de la obtención del movimiento Browniano a partir del paseo aleatorio mediante la convergencia en distribución, como se ha dicho anteriormente.

2. σ-álgebras y espacios medibles

Sean Ω un conjunto no vacío y $\mathcal{F}$ un conjunto de partes de Ω. Se dice que $\mathcal{F}$ es una *σ-álgebra* en Ω si:

(1). $\emptyset \in \mathcal{F}$.

(2). $A_n \in \mathcal{F}$, $n \in \mathbb{N}^+$, implica que $\bigcup_n A_n \in \mathcal{F}$.

(3). Si A es un elemento de $\mathcal{F}$, $(A^c =)\Omega \setminus A$, también es un elemento de $\mathcal{F}$.

Si la condición **(2)** se sustituye por:
(2'). $A_i \in \mathcal{F}$, $i \in I$ con I finito, implica que $\bigcup_{i \in I} A_i \in \mathcal{F}$,
se dice, en este caso, que $\mathcal{F}$ es un *álgebra* en Ω, (es claro que toda σ-álgebra es un álgebra).
Por las leyes de De Morgan para subconjuntos de un conjunto, se deduce que si A_n, $n = 1, 2, 3, ...$, es una sucesión infinita de elementos de una σ-álgebra $\mathcal{F}$ de partes de un conjunto Ω, entonces, $\bigcap_n A_n$ es un elemento de $\mathcal{F}$, y si $A, B \in \mathcal{F}$ y $\mathcal{F}$ es un álgebra de Ω, $A \cup B$, $A \cap B$ y $A \setminus B$, también son elementos de $\mathcal{F}$.
Si Ω es un conjunto no vacío, entonces se verifica que $\mathcal{F}_*(\Omega) = \{\emptyset, \Omega\}$, $\mathcal{F}_N(\Omega) = \{A : A \subset \Omega, \text{ y } A \text{ o } A^c \text{ es numerable}\}$ y $\mathcal{F}^*(\Omega) = \{A : A \subset \Omega\}$ son σ-álgebras sobre Ω.

Ejemplo 1 (Generación de σ-álgebras). Sea $\mathcal{E}$ un conjunto de partes de Ω. Entonces, el conjunto

$$\bigcap_{\substack{\mathcal{E} \subset \mathcal{F} \\ \mathcal{F}\ \sigma-\acute{a}lgebra\ de\ \Omega}} \mathcal{F}(= \sigma^{\Omega}(\mathcal{E}))$$

es una σ-álgebra sobre Ω y es la más pequeña σ-álgebra sobre Ω conteniendo a $\mathcal{E}$, (obsérvese que $\mathcal{F}^*(\Omega) = \{A : A \subset \Omega\}$ es σ-álgebra en Ω y $\mathcal{E} \subset \mathcal{F}^*(\Omega)$, lo que da consistencia a la definición-construcción precedente).
Cuando no haya lugar a confusión, el generador de σ-álgebras σ^Ω se abreviará por σ. En particular, si $\mathcal{E} = \{\emptyset\}$, $\sigma^\Omega(\mathcal{E}) = \mathcal{F}_*(\Omega)$, y si $\mathcal{E} = \{\{\omega\} : \omega \in \Omega\}$, entonces, $\sigma^\Omega(\mathcal{E}) = \mathcal{F}_N(\Omega)$.

Lema 2.1. *Sean $\mathcal{E} = \{E_i : i \in I\}$ un conjunto de partes de un conjunto Ω, donde I es un conjunto infinito no numerable. Entonces, si $E \in \sigma^\Omega(\mathcal{E})$, se verifica que existe un subconjunto infinito numerable, J de I tal que $E \in \sigma^\Omega(\mathcal{E}_J)$, donde $\mathcal{E}_J = \{E_i : i \in J\}$. ($\mathcal{E}_J \subset \mathcal{E}$ y $\sigma^\Omega(\mathcal{E}_J) \subset \sigma^\Omega(\mathcal{E})$).*

Lema 2.2. *Sean Ω un conjunto no vacío, $\mathcal{F}$ una σ-álgebra en Ω y $\mathcal{E}$ un conjunto de partes de Ω con $\sigma^\Omega(\mathcal{E}) = \mathcal{F}$. Sea $A \subset \Omega$ con $A \neq \emptyset$ y tal que para todo $E \in \mathcal{E}$, $A \subset E$ o $A \subset E^c$. Entonces, para todo $F \in \mathcal{F}$, se verifica que $A \subset F$ o $A \subset F^c$.*

Demostración. Sea $\mathcal{F}_1 = \{F : F \in \mathcal{F}, \text{ y } A \subset F \text{ o } A \subset F^c\}$. Se comprueba fácilmente que $\mathcal{F}_1$ es una σ-álgebra en Ω. Como $\mathcal{E} \subset \mathcal{F}_1 \subset \mathcal{F}$, se deduce que $\mathcal{F} = \sigma^\Omega(\mathcal{E}) \subset \sigma^\Omega(\mathcal{F}_1) = \mathcal{F}_1 \subset \mathcal{F}$ y $\mathcal{F}_1 = \mathcal{F}$. □

Proposición 2.3. *Sean Ω y E conjuntos, y $f : \Omega \to E$ una aplicación.*
(1). *Si $\mathcal{F}$ es una σ-álgebra en Ω y $\mathcal{E}$ es una σ-álgebra en E, se verifica que $\mathcal{E}_1 = \{A : A \in \mathcal{E} \text{ y } f^{-1}(A) \in \mathcal{F}\}$ es σ-álgebra en E contenida en $\mathcal{E}$, ($f^{-1}(\mathcal{E}_1) \subset \mathcal{F}$).*
(2). *Si $\mathcal{E}$ es una σ-álgebra en E, se tiene que $f^{-1}(\mathcal{E}) = \left\{f^{-1}(A) : A \in \mathcal{E}\right\}$ es una σ-álgebra en Ω.*
(3). *Si $\mathcal{E}^*$ es un conjunto de partes de E, se tiene: $\sigma^\Omega(f^{-1}(\mathcal{E}^*)) = f^{-1}(\sigma^E(\mathcal{E}^*))$.*
En particular, si $\Omega \subset E$ y f es la inclusión $i : \Omega \hookrightarrow E$, entonces:
(1*). *Si $\mathcal{F}$ es σ-álgebra en Ω y $\mathcal{E}$ es σ-álgebra en E, se tiene que $\mathcal{E}_1 = \{A : A \in \mathcal{E} \text{ y } A \cap \Omega \in \mathcal{F}\}$ es σ-álgebra en E contenida en $\mathcal{E}$, ($\{A \cap \Omega : A \in \mathcal{E}_1\} \subset \mathcal{F}$).*
(2*). *Si $\mathcal{E}$ es σ-álgebra en E, entonces $\mathcal{E} \cap \Omega = \{A \cap \Omega : A \in \mathcal{E}\}$ es σ-álgebra en Ω.*
(3*). *Si $\mathcal{E}^*$ es un conjunto de partes de E, se tiene: $\sigma^\Omega(\mathcal{E}^* \cap \Omega) = \sigma^E(\mathcal{E}^*) \cap \Omega$.*

Al par $(\Omega, \mathcal{F})$, se le llama *espacio medible* si:
(a). Ω es un conjunto no vacío,

(b). $\mathcal{F}$ es una σ-álgebra de subconjuntos de Ω.

Sean $(\Omega, \mathcal{F})$, $(E, \mathcal{E})$ espacios medibles y $f : \Omega \to E$ una aplicación. Se dice que f es $\mathcal{F}|\mathcal{E}$-medible o que f es una *aplicación medible* o *variable aleatoria* de $(\Omega, \mathcal{F})$ en $(E, \mathcal{E})$, si $f^{-1}(A) \in \mathcal{F}$ para todo $A \in \mathcal{E}$.

Observación 1. Es claro que si f es una aplicación medible de $(\Omega, \mathcal{F})$ en $(E, \mathcal{E})$ y g es una aplicación medible de $(E, \mathcal{E})$ en $(G, \mathcal{G})$, entonces $g \circ f$ es una aplicación medible de $(\Omega, \mathcal{F})$ en $(G, \mathcal{G})$. Además, $1_\Omega : \Omega \to \Omega$, $\omega \mapsto \omega$, es una aplicación $\mathcal{F}|\mathcal{F}$-medible, y toda aplicación constante c_{e_0}, $e_0 \in E$, de Ω en E, $\omega \mapsto e_0$, es $\mathcal{F}|\mathcal{E}$-medible.

El test, en la definición anterior, de función $\mathcal{F}|\mathcal{E}$-medible extendido a todo $A \in \mathcal{E}$ se puede reducir,

Lema 2.4. *Sean $(\Omega, \mathcal{F})$, $(E, \mathcal{E})$ espacios medibles, $f : \Omega \to E$ una función y $\mathcal{G}$ un conjunto de partes de E tal que $\sigma^E(\mathcal{G}) = \mathcal{E}$. Entonces, f es $\mathcal{F}|\mathcal{E}$-medible si y sólo si $f^{-1}(G) \in \mathcal{F}$ para todo $G \in \mathcal{G}$.*

El producto directo de espacios medibles se introduce de la siguiente forma:
Sean $(\Omega_s, \mathcal{F}_s)$, $s \in S$, espacios medibles. Sea $\Omega = \prod_{s\in S} \Omega_s$ el producto cartesiano de los conjuntos Ω_s, $s \in S$. Para cada $s \in S$ denotamos por p_s a la proyección natural de Ω sobre Ω_s, ($p_s(\omega) = \omega_s$, $\omega \in \Omega$). Se considera el conjunto de partes de Ω: $\mathcal{A} = \left\{p_s^{-1}(F) : F \in \mathcal{F}_s,\ s \in S\right\}$. Entonces tenemos la σ-álgebra $\sigma^\Omega(\mathcal{A})$ en Ω, que se designa también por $\bigotimes_{s\in S} \mathcal{F}_s$, y el espacio medible

$$\left(\bigotimes_{s\in S} (\Omega_s, \mathcal{F}_s) = \right) \left(\Omega, \bigotimes_{s\in S} \mathcal{F}_s\right),$$

que se llama *producto directo* de la familia $(\Omega_s, \mathcal{F}_s)$, $s \in S$. Es claro que $\sigma^\Omega(\mathcal{A})$ es la más pequeña σ-álgebra de Ω que hace medibles las proyecciones p_s, $s \in S$.

Proposición 2.5. *Sea $(\Omega_s, \mathcal{F}_s)$, $s \in S$, una familia de espacios medibles. Se considera $\Omega = \prod_{s\in S} \Omega_s$. Para cada $s \in S$ sea $\mathcal{E}_s$ un conjunto de subconjuntos de Ω_s tal que $\sigma^{\Omega_s}(\mathcal{E}_s) = \mathcal{F}_s$. Entonces $\mathcal{A}_1 = \left\{p_s^{-1}(E) : E \in \mathcal{E}_s,\ s \in S\right\}$, cumple que $\sigma^\Omega(\mathcal{A}_1) = \bigotimes_{s\in S} \mathcal{F}_s = \sigma^\Omega(\mathcal{A})$.*

El resultado que sigue da una caracterización importante y útil de la σ-álgebra $\bigotimes_{s\in S} \mathcal{F}_s$.

Proposición 2.6. *Sean $(\Omega_s, \mathcal{F}_s)$, $s \in S$, una familia de espacios medibles y $\mathcal{G}$ una σ-álgebra en $\Omega = \prod_{s\in S} \Omega_s$. Entonces, $\mathcal{G} = \bigotimes_{s\in S} \mathcal{F}_s$ si y sólo si:*

1. *Para cada $s \in S$, p_s es $\mathcal{G}|\mathcal{F}_s$-medible.*
2. *Para cada espacio medible $(E, \mathcal{E})$ y cada aplicación $f : E \to \Omega$, se tiene: f es $\mathcal{E}|\mathcal{G}$-medible si y sólo si $p_s \circ f$ es $\mathcal{E}|\mathcal{F}_s$-medible para todo $s \in S$.*

3. σ-álgebra de Borel y σ-álgebra de Baire de un espacio topológico

Sea $(X, \mathcal{T})$ un espacio topológico. A la σ-álgebra en X, $\sigma^X(\mathcal{T})$, se le llama *σ-álgebra de Borel* de $(X, \mathcal{T})$ y se designa también por $\mathcal{B}(X, \mathcal{T})$ que se abrevia poniendo $\mathcal{B}(X)$ cuando no hay confusión sobre la topología que se considera en X. A los elementos de $\mathcal{B}(X, \mathcal{T})$ se les llama *conjuntos de Borel* de $(X, \mathcal{T})$.
Si $\mathcal{T}_u$ es la topología usual de $\mathbb{R}$, de acuerdo con las notaciones introducidas, a la σ-álgebra de Borel de $(\mathbb{R}, \mathcal{T}_u)$ se le designará por $\mathcal{B}(\mathbb{R})$.

Observación 2. (1). Si $\mathcal{C}_\mathcal{T}$ es la familia de cerrados de $(X,\mathcal{T})$, entonces $\sigma^X(\mathcal{C}_\mathcal{T}) = \sigma^X(\mathcal{T}) = \mathcal{B}(X)$, ya que $\mathcal{C}_\mathcal{T} \subset \sigma^X(\mathcal{T})$ y $\mathcal{T} \subset \sigma^X(\mathcal{C}_\mathcal{T})$.

(2). Si $(X,\mathcal{T})$ es un espacio topológico de Hausdorff, entonces $\mathcal{K}_\mathcal{T} = \{K : K \subset X$ y K es compacto en $(X,\mathcal{T})\} \subset \mathcal{B}(X)$, (con esta hipótesis, todo compacto es cerrado). Para esta inclusión la condición de Hausdorff es esencial, como lo prueba el ejemplo que sigue: Sea $(X,\mathcal{T}) = (\mathbb{R}, \mathcal{T}_{CF})$, donde $\mathcal{T}_{CF} = \{A : A \subset \mathbb{R}$ y A^c es finito$\} \cup \{\emptyset\}$. En este caso, se tiene que $\mathcal{K}_{\mathcal{T}_{CF}} = \mathcal{F}^*(\mathbb{R})$, y sin embargo, $\sigma^\mathbb{R}(\mathcal{T}_{CF}) = \mathcal{F}_N(\mathbb{R})$, (véanse los casos particulares del **Ejemplo 1**).

Otra σ-álgebra importante que se considera en los espacios topológicos es la denominada *σ-álgebra de Baire* que se introduce a continuación.
Sea $(X,\mathcal{T})$ un espacio topológico. Se realiza la siguiente construcción:

$$\bigcap\{\mathcal{A} : \ \mathcal{A} \text{ es } \sigma - \text{álgebra en } X \text{ y } \mathcal{A} \text{ hace medibles de } \ (X,\mathcal{A}) \ \text{ en } (\mathbb{R},\mathcal{B}(\mathbb{R}))$$
$$\text{a todas las funciones continuas de } (X,\mathcal{T}) \text{ en } (\mathbb{R},\mathcal{T}_u)\},$$

(obsérvese que la σ-álgebra $\mathcal{F}^*(X)$ es un elemento de esta intersección).
Queda así definida sin ambigüedad una σ-álgebra en X que se designa por $\mathcal{B}_a(X,\mathcal{T})$ o bien $\mathcal{B}_a(X)$, si no hay confusión sobre la topología $\mathcal{T}$ que se considera en X, y se llama σ-álgebra de Baire de $(X,\mathcal{T})$. A los elementos de $\mathcal{B}_a(X,\mathcal{T})$ se les llama *subconjuntos de Baire* de $(X,\mathcal{T})$. Entonces:
(a). $\mathcal{B}_a(X)$ hace medibles a todas las funciones continuas de $(X,\mathcal{T})$ en $(\mathbb{R},\mathcal{T}_u)$.
(b). $\mathcal{B}_a(X)$ es la σ-álgebra en X más pequeña con esta propiedad **(a)**, es decir, si $\mathcal{A}$ es σ-álgebra en X con la propiedad **(a)**, entonces $\mathcal{B}_a(X) \subset \mathcal{A}$.
Con las notaciones introducidas anteriormente, se tiene que

$$\mathcal{B}_a(X,\mathcal{T}) = \sigma^X\{f^{-1}(A) : f \in C((X,\mathcal{T}),(\mathbb{R},\mathcal{T}_u)), A \in \mathcal{B}(\mathbb{R})\}.$$

Usualmente, si no ha lugar a confusión, el conjunto $C((X,\mathcal{T}),(\mathbb{R},\mathcal{T}_u))$ de aplicaciones continuas de $(X,\mathcal{T})$ en $(\mathbb{R},\mathcal{T}_u)$ se designa por $C(X,\mathbb{R})$.

Proposición 3.1. *Sea $(X,\mathcal{T})$ un espacio topológico. Entonces:*

(i). *Se tiene que $\mathcal{B}_a(X) \subset \mathcal{B}(X)$ y hay casos en los que el contenido es estricto.*

(ii). *Si $(X,\mathcal{T})$ es un espacio topológico perfectamente normal, (en particular, $(X,\mathcal{T})$ seudometrizable), se verifica que $\mathcal{B}_a(X) = \mathcal{B}(X)$.*

Demostración. **(i).** Para la primera parte, basta observar que $\mathcal{B}(X) = \sigma^X(\mathcal{T})$ hace medible a todas las aplicaciones continuas de $(X,\mathcal{T})$ en $(\mathbb{R},\mathcal{T}_u)$.
Para la última parte, consideramos el espacio topológico $(\mathbb{R},\mathcal{T}_{CN})$, ($\mathcal{T}_{CN}$ es la topología de los complementos de partes numerables de $\mathbb{R}$ y el conjunto vacío). Como las aplicaciones continuas de $(\mathbb{R},\mathcal{T}_{CN})$ en $(\mathbb{R},\mathcal{T}_u)$ son las aplicaciones constantes, se tiene

que $\mathcal{B}_a(\mathbb{R}, \mathcal{T}_{CN}) = \{\emptyset, \mathbb{R}\}$. Sin embargo, $\mathcal{B}(\mathbb{R}, \mathcal{T}_{CN}) = \mathcal{T}_{CN} \cup \mathcal{C}_{CN} = \mathcal{F}_N(\mathbb{R})$.
(ii). Sea C un subconjunto cerrado no vacío de $(X, \mathcal{T})$. Por la **Proposición 4.3.52** página 232 de [6], se tiene que existe $g \in C(X, \mathbb{R})$ tal que $g^{-1}(0) = C$. Por tanto, $C \in \mathcal{B}_a(X)$ y $\mathcal{B}(X) \subset \mathcal{B}_a(X)$, $(\sigma^X(\mathcal{C}_\mathcal{T}) = \mathcal{B}(X)$, $\mathcal{C}_\mathcal{T}$ familia de los conjuntos cerrados de $(X, \mathcal{T}))$, (**Observación 2 (1)**). □

Los tres resultados que siguen son importantes en la teoría de σ-álgebras (*tribus*, en la terminología de Bourbaki).

Proposición 3.2. *Sean $(X, \mathcal{T})$ un espacio topológico cuya topología tiene una base numerable, (es decir, $(X, \mathcal{T})$ es II.A.N), y $\mathcal{G}$ una base de $\mathcal{T}$. Entonces, $\sigma^X(\mathcal{G}) = \mathcal{B}(X) = \sigma^X(\mathcal{T})$, (de la hipótesis se deduce que todo elemento de $\mathcal{T}$ es unión numerable de elementos de $\mathcal{G}$, (véase el **Teorema 3.2.13.**, pág. 165 de [6]).*

En el resultado $(\sigma^X(\mathcal{G}) = \sigma^X(\mathcal{T}))$, de la proposición anterior, es esencial que $\mathcal{T}$ tenga una base numerable, ya que, por ejemplo, si $\mathcal{T}$ es la topología discreta en $\mathbb{R}$, entonces $\mathcal{G} = \{\{x\} : x \in \mathbb{R}\}$ es una base de $\mathcal{T}$ y $\sigma^\mathbb{R}(\mathcal{G}) = \{A : A \subset \mathbb{R}$, y A o $\mathbb{R} \setminus A$ es numerable$\} = \mathcal{F}_N(\mathbb{R})$ es distinta de $\sigma^\mathbb{R}(\mathcal{T}) = \mathcal{T} = \mathcal{F}^*(\mathbb{R})$ (conjunto de las partes de $\mathbb{R}$). Se observa que $\mathcal{T}$ no tiene bases numerables pues toda base suya contiene a $\mathcal{G}$.

Proposición 3.3. *Sea $(X, \mathcal{T})$ un espacio topológico. Si Y es un subconjunto de X, entonces $\sigma^Y(\mathcal{T}|_Y) = \sigma^X(\mathcal{T}) \cap Y$, (**(3*)** de la **Proposición 2.3**), donde $\mathcal{T}|_Y = \{A \cap Y : A \in \mathcal{T}\}$ es la topología relativa. Además, si $Y \in \mathcal{B}(X) = \sigma^X(\mathcal{T})$, se tiene $\mathcal{B}(Y) = \sigma^Y(\mathcal{T}|_Y) = \mathcal{B}(X) \cap Y = \{B \cap Y : B \in \mathcal{B}(X)\} = \{B : B \in \mathcal{B}(X)$ y $B \subset Y\}$.*

Proposición 3.4. *Sean $\{(\Omega_s, \mathcal{T}_s)\}_{s \in S}$ una familia de espacios topológicos y $\mathcal{T}_p$ la topología en $\prod_{s \in S} \Omega_s (= \Omega)$ del espacio topológico producto $\prod_{s \in S}(\Omega_s, \mathcal{T}_s)$. Entonces:*

(1). *Se considera (a partir de la colección dada) la familia de espacios medibles $(\Omega_s, \mathcal{F}_s)$, $\mathcal{F}_s = \sigma^{\Omega_s}(\mathcal{T}_s)$, $s \in S$, y el espacio medible producto directo de esta familia, $\bigotimes_{s \in S}(\Omega_s, \mathcal{F}_s) = (\Omega, \bigotimes_{s \in S} \mathcal{F}_s)$. Por las definiciones y resultados sobre el producto directo de espacios medibles, expuestas anteriormente, se tiene:*

$$\bigotimes_{s \in S} \mathcal{F}_s = \bigotimes_{s \in S} \sigma^{\Omega_s}(\mathcal{T}_s) = \sigma^\Omega(\mathcal{A}) = \sigma^\Omega(\mathcal{A}_{top}) = \sigma^\Omega(\mathcal{B}_{\mathcal{T}_p}) \subset \sigma^\Omega(\mathcal{T}_p), \quad \textit{donde:}$$

- $\mathcal{A} = \{p_s^{-1}(B) : B \in \mathcal{F}_s,\ s \in S\}$, $\mathcal{A}_{top} = \{p_s^{-1}(E) : E \in \mathcal{T}_s,\ s \in S\}$
- $\mathcal{A}_{top}$ *es subbase de* $\mathcal{T}_p$ *y por tanto*

$$(\mathcal{B}_{\mathcal{T}_p} =)\{\textit{intersecciones finitas de elementos de } \mathcal{A}_{top}\}$$

  es base de $\mathcal{T}_p$ *y* $\sigma^\Omega(\mathcal{B}_{\mathcal{T}_p}) = \sigma^\Omega(\mathcal{A}_{top})$.
- *Respecto al contenido $\sigma^\Omega(\mathcal{B}_{\mathcal{T}_p}) \subset \sigma^\Omega(\mathcal{T}_p)$ tenemos el ejemplo: $(\Omega_s, \mathcal{T}_s) = (\mathbb{R}, \mathcal{T}_u)$, $s \in S$, S conjunto infinito no numerable. Como la diagonal Δ es un cerrado de $(\mathbb{R}, \mathcal{T}_u)^S$, se tiene que $\Delta \in \sigma^\Omega(\mathcal{T}_p)$. Sin embargo, $\Delta \notin \sigma^\Omega(\mathcal{B}_{\mathcal{T}_p})$, (véase [4], pág. 70), y por tanto, en este caso el contenido $\sigma^\Omega(\mathcal{B}_{\mathcal{T}_p}) \subset \sigma^\Omega(\mathcal{T}_p)$ es estricto.*

(2). *Supongamos que S es numerable y para cada $s \in S$ la topología $\mathcal{T}_s$ tiene una base numerable, $\mathcal{B}_s$. Entonces $\bigotimes_{s\in S}\sigma^{\Omega_s}(\mathcal{T}_s) = \sigma^{\Omega}(\mathcal{T}_p)$.*
*En efecto: Por la **Proposición 3.2.7.** de [6], (pág. 164), la topología $\mathcal{T}_p$ tiene una base numerable. Así, por la **Proposición 3.2**, se concluye que:*

$$\bigotimes_{s\in S}\mathcal{F}_s = \bigotimes_{s\in S}\sigma^{\Omega_s}(\mathcal{T}_s) = \sigma^{\Omega}(\mathcal{A}) = \sigma^{\Omega}(\mathcal{A}_{top}) = \sigma^{\Omega}(\mathcal{B}_{\mathcal{T}_p}) = \sigma^{\Omega}(\mathcal{T}_p).$$

En **(2)**, de la proposición anterior, es esencial que S sea numerable, como lo prueba el ejemplo $(\Omega_s,\mathcal{T}_s) = (\mathbb{R},\mathcal{T}_u)$, $s \in S$, S conjunto infinito no numerable.
En relación con la esencialidad de las hipótesis de la **Proposición 3.4(2)** se tiene también:

Proposición 3.5. *Sea $(X,\mathcal{T})$ un espacio topológico de Hausdorff con* cardinal$(X) > 2^{\aleph_0}$*. Entonces, $\sigma^X(\mathcal{T})\otimes\sigma^X(\mathcal{T}) \subset \sigma^{X\times X}(\mathcal{T}_p)$, la diagonal*

$$\Delta = \{(x,x) : x\in X\} \text{ no pertenece a } \sigma^X(\mathcal{T})\otimes\sigma^X(\mathcal{T}) \text{ y } \Delta\in\sigma^{X\times X}(\mathcal{T}_p),$$

donde $\mathcal{T}_p$ es la topología producto de $(X,\mathcal{T})\times(X,\mathcal{T})$, (obsérvese que las hipótesis sobre el espacio topológico que se considera, implican que $\mathcal{T}$ no tiene ninguna base numerable: Si $\mathcal{T}$ tuviese una base numerable $\mathcal{B}$, como el espacio es de Hausdorff, todo punto de X se obtiene como intersección de elementos de $\mathcal{B}$, concretamente $\{x\} = \bigcap\{B : B\in\mathcal{B}, x\in B\}$. Así, la aplicación $f : X\to\mathcal{F}^*(\mathcal{B})$, $x\mapsto\{B : B\in\mathcal{B}, x\in B\}$, es inyectiva, y por tanto $cardinal(X) \leqslant cardinal(\mathcal{F}^*(\mathcal{B})) \leqslant cardinal(\mathcal{F}^*(\mathbb{N})) = 2^{\aleph_0}$). Por tanto, se tiene que $\sigma^X(\mathcal{T})\otimes\sigma^X(\mathcal{T}) \subsetneqq \sigma^{X\times X}(\mathcal{T}_p)$.

Descripción de conjuntos de Borel mediante seudométricas

En el caso de la topología $\mathcal{T}_d$ asociada a una seudométrica d en un conjunto X, es bien conocido que $\mathcal{T}_d$ tiene una base numerable si y sólo si $(X,\mathcal{T}_d)$ es separable, (es decir, existe un subconjunto numerable y denso en $(X,\mathcal{T}_d)$).
Por tanto, por la **Proposición 3.2**, si (X,d) es un espacio seudométrico separable y $\mathcal{B}_d$ es la base de $\mathcal{T}_d$ formada por todas las bolas abiertas en la seudométrica d, se verifica que $\mathcal{B}(\mathcal{T}_d) = \sigma^X(\mathcal{B}_d)$. Además, con esta misma hipótesis de separabilidad, si $\mathcal{B}_d^=$ es el conjunto de todas las bolas cerradas en la seudométrica d, también se tiene que $\mathcal{B}(\mathcal{T}_d) = \sigma^X(\mathcal{B}_d^=)$, (basta observar que $B(x;\varepsilon) = \{y\in X : d(y,x)<\varepsilon\} = \bigcup_{n\in\mathbb{N}^+} B^=(x;\varepsilon-\varepsilon/(n+1)))$ y que los elementos de $\mathcal{B}_d^=$ son cerrados de $(X,\mathcal{T}_d)$).

Como casos particulares importantes, de lo anterior, tenemos las σ-álgebras de los espacios Euclideanos. Sean las métricas: $\rho_0(x,y) = |x-y|$, métrica en $\mathbb{R}$ cuya topología asociada es $\mathcal{T}_u$; $\rho_1(x,y) = |x-y|/(1+|x-y|)$, métrica en $\mathbb{R}$ equivalente a ρ_0, (dos métricas en un mismo conjunto son equivalentes si definen la misma topología), y por tanto con topología asociada $\mathcal{T}_u$; $\rho_n(x,y) = 2^{-1}\rho_1(x_1,y_1)+...+2^{-n}\rho_1(x_n,y_n)$ métrica en $\mathbb{R}^n$ equivalente a la métrica euclídea y, por tanto, con topología asociada

$\mathcal{T}_u^n$; $\rho_\infty(x,y) = \sum_{k=1}^{k=\infty} 2^{-k}\rho_1(x_k, y_k)$, métrica en $\mathbb{R}^\infty$ con topología asociada $\mathcal{T}_u^\infty$. Estos espacios topológicos son separables y por tanto:

$$\begin{aligned} \sigma^{\mathbb{R}}(\{\text{bolas abiertas de } \rho_1\}) &= \mathcal{B}(\mathbb{R}) &= \sigma^{\mathbb{R}}(\mathcal{T}_u) \\ \sigma^{\mathbb{R}^n}(\{\text{bolas abiertas de } \rho_n\}) &= \mathcal{B}(\mathbb{R}^n) &= \sigma^{\mathbb{R}^n}(\mathcal{T}_u^n) \\ \sigma^{\mathbb{R}^\infty}(\{\text{bolas abiertas de } \rho_\infty\}) &= \mathcal{B}(\mathbb{R}^\infty) &= \sigma^{\mathbb{R}^\infty}(\mathcal{T}_u^\infty). \end{aligned}$$

En el producto cartesiano $\mathbb{R}^S$, S conjunto infinito no numerable, hemos visto que $\bigotimes_{s\in S}(\sigma^{\mathbb{R}}(\mathcal{T}_u))_s \subsetneqq \sigma^{\mathbb{R}^S}(\mathcal{T}_u^S)$, $((\sigma^{\mathbb{R}}(\mathcal{T}_u))_s = \sigma^{\mathbb{R}}(\mathcal{T}_u)$, $s \in S)$. Observamos que el espacio topológico $(\mathbb{R}^S, \mathcal{T}_u^S) = \prod_{s\in S}(\mathbb{R}, \mathcal{T}_u)_s$ no es metrizable y no tiene bases numerables, (**Proposición 1.3.44.** (pág. 54) y **Proposición 3.2.7.** (pág. 164) de [6]).

4. Probabilidades en espacios topológicos

Sea $(X, \mathcal{T})$ un espacio topológico. Una *medida de Borel finita* en $(X, \mathcal{B}(X))$, $(\mathcal{B}(X) = \sigma^X(\mathcal{T}))$, es una aplicación $\mu : \mathcal{B}(X) \to \mathbb{R}$ tal que:

(1). $\mu(\emptyset) = 0$.

(2). $\mu(B) \geqslant 0$, para todo $B \in \mathcal{B}(X)$.

(3). Si $B_1, B_2, \ldots$ es una cantidad numerable de elementos de $\mathcal{B}(X)$ disjuntos dos a dos, entonces, $\mu\left(\bigcup_{n\in\mathbb{N}^+} B_n\right) = \sum_{n\in\mathbb{N}^+} \mu(B_n)$.

Una medida de Borel finita en $(X, \mathcal{B}(X))$ es una *probabilidad* (de Borel) en $(X, \mathcal{B}(X))$ si además $\mu(X) = 1$.

Ejemplos de probabilidades en $(X, \mathcal{B}(X))$ son las probabilidades de Dirac: Para todo $x \in X$, $\delta_x : \mathcal{B}(X) \to \{0, 1\}$, dada por $\delta_x(B) = 0$ si $x \notin B$ y $\delta_x(B) = 1$ si $x \in B$, $B \in \mathcal{B}(X)$, es una probabilidad en $(X, \mathcal{B}(X))$ que se llama *probabilidad de Dirac en el punto x*.

Lema 4.1. *Sean $(X, \mathcal{T})$ un espacio topológico, μ una medida de Borel finita en $(X, \mathcal{B}(X))$ y $\{B_i\}_{i\in I}$ un conjunto de elementos de $\mathcal{B}(X)$ disjuntos dos a dos. Entonces, $J = \{i \in I : \mu(B_i) > 0\}$ es un conjunto numerable.*

Demostración. Para cada $n \in \mathbb{N}^+$, sea $I_n = \left\{i : i \in I \text{ y } \mu(B_i) > \frac{1}{n}\right\} \subset J$. Como existe $m \in \mathbb{N}^+$ tal que $\frac{m}{n} > \mu(X)$, es claro que I_n no puede tener más de m elementos. Así, $J = \bigcup_{n\in\mathbb{N}^+} I_n$ es numerable. □

Definición 4.2. *Un espacio topológico $(X, \mathcal{T})$ que es seudometrizable por una seudométrica completa y separable, es decir, existe una seudométrica d en X tal que $\mathcal{T} = \mathcal{T}_d$ (topología asociada a la seudométrica d) y (X, d) es un espacio seudométrico completo separable, se le llama* espacio topológico **s**-polaco.

Si en la definición anterior la seudométrica d es una métrica, se tienen los *espacios topológicos polacos.* Observamos que los espacios $(\mathbb{R}^n, \mathcal{T}_u^n)$, $n \in \mathbb{N}^+$, son espacios topológicos polacos. También es polaco el espacio topológico $(\mathbb{R}^\infty, \mathcal{T}_u^\infty)$, (véase **Proposición 1.3.43.**, (pág. 54), y **Proposición 3.4.8.**, (pág. 173), de [6], (la métrica utilizada en la demostración de la **Proposición 1.3.43.** es completa)).
Por otro lado, $(\mathbb{R}, \mathcal{T}_t)$, donde $\mathcal{T}_t = \{\emptyset, \mathbb{R}\}$ (topología trivial), es un espacio **s**-polaco que no es espacio topológico polaco. De hecho, un espacio topológico **s**-polaco es polaco si y sólo si es de Hausdorff.

Proposición 4.3 (**Propiedad de regularidad**). *Sean $(X, \mathcal{T})$ un espacio topológico y P una probabilidad en el espacio medible $(X, \mathcal{B}(X))$. Supongamos que cada subconjunto cerrado de $(X, \mathcal{T})$ es G_δ, (un espacio seudometrizable tiene esta propiedad). Entonces, para todo $B \in \mathcal{B}(X)$,*

$$\begin{aligned} P(B) &= \sup\{P(C) : \ C \subset B, \ C \textit{ cerrado en } (X, \mathcal{T})\} = \\ &= \inf\{P(A) : \ B \subset A, \ A \in \mathcal{T}\}, \ \textit{(en un planteamiento general la} \\ &\qquad \textit{medida interior es menor o igual que la medida exterior).} \end{aligned}$$

Demostración. Vamos a utilizar la técnica de los buenos conjuntos. Sea $\mathcal{B}^*$ el conjunto de los $B^* \in \mathcal{B}(X)$ tales que

$$\begin{aligned} P(B^*) = \sup\{P(C) : \ C \subset B^*, \ C \text{ cerrado en } (X, \mathcal{T})\} = \\ = \inf\{P(A) : \ B^* \subset A, \ A \in \mathcal{T}\}. \end{aligned}$$

El conjunto $\mathcal{B}^*$ contiene a todos los subconjuntos cerrados de $(X, \mathcal{T})$. En efecto: Sea C un cerrado de $(X, \mathcal{T})$. Por la hipótesis G_δ, existen G_1, G_2,... elementos de $\mathcal{T}$, tales que $C = \bigcap_{n \in \mathbb{N}^+} G_n$. Se consideran los abiertos $A_n = G_1 \cap ... \cap G_n$, $n \in \mathbb{N}^+$. Estos abiertos cumplen que $A_{n+1} \subset A_n$, $n \in \mathbb{N}^+$, y $C = \bigcap_{n \in \mathbb{N}^+} A_n$. Así, por el **Teorema 3.1.11.(II)(1)**, (pág. 12), de [4], tomando $(X, \mathcal{B}(X), P)$, se tiene que $P(C) = \lim_{n \to +\infty} P(A_n)$, lo cual prueba que $C \in \mathcal{B}^*$, ya que $P(C) = \inf\{P(A) : C \subset A, \ A \in \mathcal{T}\}$.
El conjunto $\mathcal{B}^*$ es una σ-álgebra sobre X. En efecto:
(1). Es evidente que $\emptyset$ (abierto y cerrado) es un elemento de $\mathcal{B}^*$.
(2). Sean $B \in \mathcal{B}^*$ y $\varepsilon > 0$. Entonces, existen un cerrado C_ε y un abierto A_ε tales que $C_\varepsilon \subset B \subset A_\varepsilon$, $P(B) - \varepsilon < P(C_\varepsilon)$ y $P(B) + \varepsilon > P(A_\varepsilon)$. Así,

$$\begin{aligned} (A_\varepsilon)^c \subset B^c \subset (C_\varepsilon)^c, \ \ P(B^c) = 1 - P(B) > 1 - P(C_\varepsilon) - \varepsilon = P((C_\varepsilon)^c) - \varepsilon \ \text{ y} \\ P(B^c) = 1 - P(B) < 1 - P(A_\varepsilon) + \varepsilon = P((A_\varepsilon)^c) + \varepsilon. \end{aligned}$$

Como $(C_\varepsilon)^c$ es abierto y $(A_\varepsilon)^c$ es cerrado, se concluye que $B^c \in \mathcal{B}^*$.
(3). Sean B_1, B_2,... elementos de $\mathcal{B}^*$ y $\varepsilon > 0$. Entonces, para cada $n \in \mathbb{N}^+$, existen un cerrado C_n y un abierto A_n tales que $C_n \subset B_n \subset A_n$, $P(B_n) - \varepsilon/2^{n+1} < P(C_n)$

y $P(A_n) < P(B_n) + \varepsilon/2^n$.
Se tiene que $\bigcup_{n\in\mathbb{N}^+} B_n \subset \bigcup_{n\in\mathbb{N}^+} A_n$, $\bigcup_{n\in\mathbb{N}^+} A_n$ es abierto y

$$P\left(\bigcup_{n\in\mathbb{N}^+} A_n\right) - P\left(\bigcup_{n\in\mathbb{N}^+} B_n\right) = P\left(\bigcup_{n\in\mathbb{N}^+} A_n \setminus \bigcup_{n\in\mathbb{N}^+} B_n\right) \leqslant$$
$$\leqslant P\left(\bigcup_{n\in\mathbb{N}^+} (A_n \setminus B_n)\right) \leqslant \sum_{n\in\mathbb{N}^+} P(A_n \setminus B_n) < \sum_{n\in\mathbb{N}^+} \frac{\varepsilon}{2^n} = \varepsilon.$$

Por otro lado por el **Teorema 3.1.10.**, (pág. 11), de [4], se tiene que $P(\bigcup_{n\in\mathbb{N}^+} C_n) = \lim_{m\to+\infty} P(\bigcup_{n=1}^{m} C_n)$ y por tanto existe un número natural m_0 para el cual (y siguientes) $P(\bigcup_{n\in\mathbb{N}^+} C_n) - P(\bigcup_{n=1}^{m_0} C_n) < \varepsilon/2$. Entonces, $\bigcup_{n=1}^{m_0} C_n$ es un cerrado contenido en $\bigcup_{n\in\mathbb{N}^+} B_n$ y por un cálculo análogo al realizado anteriormente,

$$P\left(\bigcup_{n\in\mathbb{N}^+} B_n\right) - P\left(\bigcup_{n=1}^{m_0} C_n\right) < P\left(\bigcup_{n\in\mathbb{N}^+} B_n\right) - P\left(\bigcup_{n\in\mathbb{N}^+} C_n\right) + \frac{\varepsilon}{2} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Se concluye que $\bigcup_{n\in\mathbb{N}^+} B_n \in \mathcal{B}^*$, y se termina la demostración de que $\mathcal{B}^*$ es una σ-álgebra en X.
Puesto que $\mathcal{B}^*$ es una σ-álgebra con $\mathcal{C}_\mathcal{T} \subset \mathcal{B}^* \subset \mathcal{B}(X)$, donde $\mathcal{C}_\mathcal{T}$ es la familia de cerrados de $(X, \mathcal{T})$, $\sigma^X(\mathcal{C}_\mathcal{T}) = \mathcal{B}(X) \subset \mathcal{B}^* \subset \mathcal{B}(X)$ y $\mathcal{B}^* = \mathcal{B}(X)$. □

Proposición 4.4 (**Propiedad *tight***). *Sean $(X, \mathcal{T})$ un espacio topológico **s**-polaco y P una probabilidad en el espacio medible $(X, \mathcal{B}(X))$. Entonces para todo $\varepsilon > 0$ existe un subconjunto compacto y cerrado K_ε en $(X, \mathcal{T})$ tal que $P(K_\varepsilon) \geqslant 1 - \varepsilon$.*

Demostración. Sean d seudométrica completa en X con $\mathcal{T}_d = \mathcal{T}$, $D = \{x_n\}_{n\in N^+}$ un subconjunto numerable y denso en $(X, \mathcal{T})$ y $\varepsilon > 0$. Para cada $x \in X$ y cada número real positivo r, designamos por $B(x; r)$ al subconjunto de X, $\{y \in X : d(y, x) < r\}$ (bola abierta en la seudométrica d de centro x y radio r).
Dado $n \in \mathbb{N}^+$ se tiene que $\bigcup_{m\in\mathbb{N}^+} B(x_m; 1/n) = X$, y por tanto por el **Teorema 3.1.10.**, (pág. 11), de [4], obtenemos que

$$\lim_{m\to+\infty} P\left(B\left(x_1; \frac{1}{n}\right) \cup ... \cup B\left(x_m; \frac{1}{n}\right)\right) = 1$$

e interpretando el límite se concluye que para un m_n, y de ahí en adelante, se verifica

$$P\left(\bigcup_{k=1}^{k=m_n} B\left(x_k; \frac{1}{n}\right)\right) \geqslant 1 - \frac{\varepsilon}{2^n}, \text{ y se puede tomar } m_1 < m_2 < m_3 < ... .$$

Es claro que el conjunto $M = \bigcap_{n\in\mathbb{N}^+} \left(\bigcup_{k=1}^{k=m_n} B\left(x_k; \frac{1}{n}\right)\right)$ es precompacto

(o totalmente acotado) en el espacio seudométrico completo (X, d). Así, la adherencia K_ε de M en $(X, \mathcal{T})$ es precompacto y completo, y por tanto K_ε es compacto y cerrado en $(X, \mathcal{T})$, (**Corolario VIII.1.60**,(pág. 165), de [5]).
Por último, de nuevo por el **Teorema 3.1.10.**, (pág. 11), de [4], se deduce que

$$
\begin{aligned}
P(X \setminus K_\varepsilon) \leqslant P(X \setminus M) = P\left(\bigcup_{n\in\mathbb{N}^+}\left(X \setminus \bigcup_{k=1}^{k=m_n} B\left(x_k; \frac{1}{n}\right)\right)\right) = \\
= \lim_{n\to+\infty} P\left(\bigcup_{p=1}^{n}\left(X \setminus \bigcup_{k=1}^{k=m_p} B\left(x_k; \frac{1}{p}\right)\right)\right) \leqslant \lim_{n\to+\infty}\left(\sum_{p=1}^{n}\frac{\varepsilon}{2^p}\right) = \sum_{n=1}^{\infty}\frac{\varepsilon}{2^n} = \varepsilon,
\end{aligned}
$$

y por tanto, $P(K_\varepsilon) = 1 - P(X \setminus K_\varepsilon) \geqslant 1 - \varepsilon$. □

Corolario 4.5. *Sean $(X, \mathcal{T})$ un espacio topológico **s**-polaco, P una probabilidad en el espacio medible $(X, \mathcal{B}(X))$ y $B \in \mathcal{B}(X) = \sigma^X(\mathcal{T})$. Entonces, $P(B) = \sup\{P(K) : K \subset B,\ K$ compacto y cerrado en $(X, \mathcal{T})\}$.*

Demostración. Para cada $\varepsilon > 0$, por la proposición anterior, existe un compacto cerrado K_ε en $(X, \mathcal{T})$ tal que $P(K_\varepsilon) \geqslant 1-\varepsilon$, y por tanto $P(X \setminus K_\varepsilon) = 1-P(K_\varepsilon) \leqslant \varepsilon$. Así, por la **Propiedad de Regularidad**, aplicada al conjunto $B \cap K_\varepsilon$,

$$
\begin{aligned}
P(B) - \varepsilon \leqslant P(B) - P(X \setminus K_\varepsilon) \leqslant P(B \setminus (X \setminus K_\varepsilon)) = \\
= P(B \cap K_\varepsilon) = \sup\{P(C) :\ C \subset B \cap K_\varepsilon, C \text{ cerrado en } (X, \mathcal{T})\} \leqslant \\
\leqslant \sup\{P(K) :\ K \subset B,\ K \text{ compacto y cerrado en } (X, \mathcal{T})\} \leqslant P(B),
\end{aligned}
$$

ya que $C \subset K_\varepsilon$ y C cerrado implica que C es compacto. Como ε es arbitrario, se concluye que $P(B) = \sup\{P(K) :\ K \subset B,\ K$ compacto y cerrado en $(X, \mathcal{T})\}$. □

Teorema 4.6. *Sean $(\Omega_n, \mathcal{T}_n, \mathcal{F}_n)$, $n \in \mathbb{N}^+$, espacios medibles **s**-polacos, es decir, $(\Omega_n, \mathcal{T}_n)$ es un espacio topológico **s**-polaco y $\mathcal{F}_n$ es la σ-álgebra en Ω_n generada por la topología $\mathcal{T}_n$, $\mathcal{F}_n = \sigma^{\Omega_n}(\mathcal{T}_n)$, $n \in \mathbb{N}^+$. Se considera el producto directo de la familia de espacios medibles $(\Omega_n, \mathcal{F}_n)$, $n = 1, 2, ...$, dada:*

$$
\left(\prod_{n\in\mathbb{N}^+}\Omega_n (= \Omega), \left(\bigotimes_{n\in\mathbb{N}^+}\mathcal{F}_n = \right)\sigma^\Omega(\mathcal{A})\right), \textit{donde } \mathcal{A} = \left\{p_n^{-1}(F) :\ F \in \mathcal{F}_n,\ n \in \mathbb{N}^+\right\},
$$
$$
p_n : \Omega \to \Omega_n,\ p_n(\omega) = \omega(n) = \omega_n,\ \ \omega \in \Omega,\ \omega_n \in \Omega_n,\ n \in \mathbb{N}^+.
$$

*Se considera también Ω con la topología producto $\mathcal{T}_p$, $(\Omega, \mathcal{T}_p)$; sabemos que $\sigma^\Omega(\mathcal{T}_p) = \bigotimes_{n\in\mathbb{N}^+}\mathcal{F}_n$, (**Proposición 3.4(2)**).*
Sean P_1, P_2,... probabilidades en $(\Omega_1, \mathcal{F}_1)$, $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \bigotimes \mathcal{F}_2)$,..., respectivamente.

Supongamos que se cumple:

$$P_{n+1}(B \times \Omega_{n+1}) = P_n(B), \quad B \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n,$$
$$(\textit{por tanto, } B \times \Omega_{n+1} \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_{n+1}),\ n \in \mathbb{N}^+.$$

Entonces, existe una única probabilidad P en $\left(\prod_{n\in\mathbb{N}^+} \Omega_n, \bigotimes_{n\in\mathbb{N}^+} \mathcal{F}_n\right)$ *tal que para cada elemento $B \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n$, (por consiguiente, se tiene que $\{\omega \in \Omega : (\omega_1, ..., \omega_n) \in B\} = (p_1, ..., p_n)^{-1}(B)(= J_n(B))$ es un elemento de $\bigotimes_{n\in\mathbb{N}^+} \mathcal{F}_n$), se verifica que $P(J_n(B)) = P_n(B)$, $n \in \mathbb{N}^+$.*

Demostración. Se considera $\mathcal{J} = \{J_n(B^n) : n \in \mathbb{N}^+,\ B^n \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n\}$. Se tiene que $\mathcal{J}$ es un álgebra de partes de Ω tal que $\mathcal{A} \subset \mathcal{J}$ y $\sigma^{\Omega}(\mathcal{J}) = \bigotimes_{n\in\mathbb{N}^+} \mathcal{F}_n$.
Supongamos que $J_n(B^n) = J_{n+k}(B^{n+k})$, $B^n \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n$, $B^{n+k} \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_{n+k}$. De esta igualdad obtenemos: Si $(x_1, ..., x_{n+k}) \in \Omega_1 \times ... \times \Omega_{n+k}$, entonces: $(x_1, ..., x_n) \in B^n$ si y sólo si $(x_1, ..., x_{n+k}) \in B^{n+k}$. Por consiguiente, del análisis precedente tenemos:

$$P_n(B^n) = P_{n+1}(B^n \times \Omega_{n+1}) = P_{n+1}(\{(x_1, ..., x_{n+1}) : (x_1, ..., x_n) \in B^n\}) = ... =$$
$$= P_{n+k}(\{(x_1, ..., x_{n+k}) : (x_1, ..., x_n) \in B^n\}) = P_{n+k}(B^{n+k}).$$

Luego P está bien definida en $\mathcal{J}$ como función de conjunto (para cada $\widehat{B} \in \mathcal{J}$, $P(\widehat{B}) = P_n(B^n)$, donde $B^n \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n$ y $\widehat{B} = J_n(B^n)$), P es mayor o igual que cero en $\mathcal{J}$ y $P(\Omega) = 1$.
La aplicación P es finitamente aditiva en $\mathcal{J}$. En efecto: Sean $\widehat{A}$ y $\widehat{C}$ dos elementos disjuntos de $\mathcal{J}$. Se tiene que $\widehat{A} = J_n(A^n)$ con $A^n \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n$ y $\widehat{C} = J_{n'}(C^{n'})$ con $C^{n'} \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_{n'}$. Sean $m = n + n'$, y

$$A^m = A^n \times \Omega_{n+1} \times ... \times \Omega_m \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_m \ \text{ y}$$
$$C^m = C^{n'} \times \Omega_{n'+1} \times ... \times \Omega_m \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_m.$$

Entonces, $\widehat{A} = J_m(A^m)$, $\widehat{C} = J_m(C^m)$, A^m y C^m son disjuntos ya que $\widehat{A}$ y $\widehat{C}$ lo son, y además $\widehat{A} + \widehat{C} = J_m(A^m + C^m)$. Así,

$$P(\widehat{A}) + P(\widehat{C}) = P(J_m(A^m)) + P(J_m(C^m)) =$$
$$= P_m(A^m) + P_m(C^m) = P_m(A^m + C^m) = P(\widehat{A} + \widehat{C}),$$

lo que prueba que P es finitamente aditiva en el álgebra $\mathcal{J}$.

Veamos que:

$$\widehat{B}_n \downarrow \emptyset,\ n \to \infty,\ \widehat{B}_n \in \mathcal{J},\ n \in \mathbb{N}^+, \text{ implica que } \lim_{n\to+\infty} P(\widehat{B}_n) = 0. \tag{4.1}$$

Con lo cual P será σ-aditiva en $\mathcal{J}$ por el **Teorema 3.1.11.(I)**, (pág. 12), de [4].
A cada $n \in \mathbb{N}^+$ se le asigna

$$k(n) = \text{mín}\left\{m \in \mathbb{N}^+ : \text{ existe } B_n^m \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_m \text{ con } J_m(B_n^m) = \widehat{B}_n\right\},$$

(Obsérvese que $k(n) = \text{máx}\{p \in \mathbb{N}^+ : \{\omega_p : \omega \in \widehat{B}_n\} \neq \Omega_p\}$). De esta definición de $k(n)$ se deduce: Si $n < n'$, entonces

$$\widehat{B}_n = J_{k(n)}\left(B_n^{k(n)}\right) \supset \widehat{B}_{n'} = J_{k(n')}\left(B_{n'}^{k(n')}\right),$$
$$B_n^{k(n)} \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_{k(n)},\ B_{n'}^{k(n')} \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_{k(n')} \text{ y } k(n) \leqslant k(n').$$

Si ocurre que la sucesión creciente de números naturales $\{k(n)\}_{n\in\mathbb{N}^+}$ está acotada superiormente, se tiene que existe $p \in \mathbb{N}^+$ tal que $k(n) = k(p)$ para todo $n > p$. Entonces:
Para todo $n > p$ existe $B_n^{k(p)} \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_{k(p)}$, $(k(n) = k(p))$, tal que $\widehat{B}_n = J_{k(p)}\left(B_n^{k(p)}\right)$. De la hipótesis $\widehat{B}_n \downarrow \emptyset$, $n \to +\infty$, se deduce que en el espacio de probabilidad $(\Omega_1 \times ... \times \Omega_{k(p)}, \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_{k(p)}, P_{k(p)})$, $B_n^{k(p)} \downarrow \emptyset$, $n \to +\infty$, y por tanto se verifica que $\text{lím}_{n\to+\infty} P_{k(p)}\left(B_n^{k(p)}\right) = 0$ lo cual implica que $\text{lím}_{n\to+\infty} P(\widehat{B}_n) = 0$, ya que

$$P(\widehat{B}_n) = P_{k(p)}\left(B_n^{k(p)}\right),\ \ \widehat{B}_n = J_{k(p)}\left(B_n^{k(p)}\right),\ \text{ para todo } n > p.$$

Luego, en esta situación, se cumple 4.1.

Supongamos ahora que la sucesión creciente de números naturales $\{k(n)\}_{n\in\mathbb{N}^+}$ no está acotada superiormente, entonces se construye una sucesión estrictamente creciente de números naturales $\{\bar{k}(n)\}_{n\in\mathbb{N}^+}$ tomando $\bar{k}(1) = k(1)$ y para todo $n > 1$, $\bar{k}(n)$ igual al primer $k(n')$ tal que $\bar{k}(n-1) < k(n')$. Es claro que $\widehat{B}_{\bar{k}(n)} \downarrow \emptyset$, $n \to +\infty$. Si probamos que

$$\lim_{n\to+\infty} P(\widehat{B}_{\bar{k}(n)}) = 0,$$

quedará probado 4.1 ya que la sucesión $P(\widehat{B}_n)$, $n \in \mathbb{N}^+$ es decreciente.
Luego se puede replantear el problema diciendo que se trata de probar la siguiente implicación:

$$\widehat{B}_n \downarrow \emptyset,\ n \to +\infty,\ \text{donde } \widehat{B}_n \in \mathcal{J},\ n \in \mathbb{N}^+,\ \widehat{B}_n = \{\omega \in \Omega : (\omega_1, ..., \omega_n) \in B_n\},$$
$$B_n \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n,\ \ (\widehat{B}_n = J_n(B_n)),\ \text{ implica que } \lim_{n\to+\infty} P(\widehat{B}_n) = 0. \quad (4.2)$$

Supongamos lo contrario, es decir, $\text{lím}_{n\to+\infty} P(\widehat{B}_n) = \delta > 0$.
Para cada $n \in \mathbb{N}^+$ nos fijamos en $(\Omega_1 \times ... \times \Omega_n, \mathcal{T}_1 \times ... \times \mathcal{T}_n = \mathcal{T}_{1,...,n}^p)$ espacio

topológico **s**-polaco (véanse las proposiciones 3.4.8. (pág. 173) y 1.3.43. (pág. 54) de [6]), $\left(\Omega_1 \times ... \times \Omega_n, \sigma\left(\mathcal{T}^p_{1,...,n}\right)\right)$, de hecho se tiene la igualdad $\sigma^{\Omega_1}(\mathcal{T}_1)\otimes ... \otimes\sigma^{\Omega_n}(\mathcal{T}_n) = \sigma\left(\mathcal{T}^p_{1,...,n}\right)$, (**Proposición 3.4(2)**), $\sigma^{\Omega_i}(\mathcal{T}_i) = \mathcal{F}_i$, $i = 1, ..., n$, $(\Omega_1 \times ... \times \Omega_n, \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n)$ y P_n probabilidad (de la hipótesis) en $(\Omega_1 \times ... \times \Omega_n, \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n)$ y $B_n \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n$ dada en 4.2. Por el **Corolario 4.5**,

$$\begin{aligned} &P_n(B_n) = \\ &\quad = \sup\left\{P_n(K) :\ K \subset B_n,\ K \text{ compacto y cerrado en } \left(\Omega_1 \times ... \times \Omega_n, \mathcal{T}^p_{1,...,n}\right)\right\}, \end{aligned}$$

(obsérvese que cada compacto y cerrado K de $\left(\Omega_1 \times ... \times \Omega_n, \mathcal{T}^p_{1,...,n}\right)$ es un elemento de $\mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n$). Luego (teniendo en cuenta el supremo) existe un subconjunto compacto y cerrado K_n en el espacio topológico $\left(\Omega_1 \times ... \times \Omega_n, \mathcal{T}^p_{1,...,n}\right)$, (por tanto, elemento de $\mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n$), tal que $K_n \subset B_n$ y $P_n(B_n \setminus K_n) \leqslant \delta/2^{n+1}$. Si ponemos

$$(\widehat{K}_n =) \{\omega \in \Omega :\ (\omega_1, ..., \omega_n) \in K_n\} = J_n(K_n) = (p_1, ..., p_n)^{-1}(K_n),$$

tenemos que $\widehat{K}_n$ es un elemento del álgebra $\mathcal{J}$ que está contenida en $\bigotimes_{n\in\mathbb{N}^+} \mathcal{F}_n$, $\widehat{K}_n \subset \widehat{B}_n$, $\bigcap_{i=1}^n \widehat{K}_i = \bigcap_{i=1}^n (p_1, ..., p_n)^{-1}(K_i)$ y

$$P\left(\widehat{B}_n \setminus \widehat{K}_n\right) = P\left(J_n\left(B_n \setminus K_n\right)\right) = P_n\left(B_n \setminus K_n\right) \leqslant \frac{\delta}{2^{n+1}}.$$

Se tienen las siguientes fórmulas:

$$\begin{aligned} &\bigcap_{k=1}^n \widehat{K}_k (= \widehat{C}_n) = J_1(K_1) \cap J_2(K_2) \cap ... \cap J_n(K_n) = \\ &\qquad J_n(K_1 \times \Omega_2 \times ... \times \Omega_n) \cap J_n(K_2 \times \Omega_3 \times ... \times \Omega_n) \cap ... \cap J_n(K_n) = \\ &\qquad = J_n\left((K_1 \times \Omega_2 \times ... \times \Omega_n) \cap (K_2 \times \Omega_3 \times ... \times \Omega_n) \cap ... \cap K_n\right), \\ &(C_n =)(K_1 \times \Omega_2 \times ... \times \Omega_n) \cap (K_2 \times \Omega_3 \times ... \times \Omega_n) \cap ... \cap K_n \in \mathcal{F}_1 \otimes ... \otimes \mathcal{F}_n, \\ &\qquad J_n(C_n) = \widehat{C}_n,\ \widehat{C}_n = \{\omega \in \Omega : (\omega_1, ..., \omega_n) \in C_n\} = (p_1, ..., p_n)^{-1}(C_n) \end{aligned}$$

y es claro que $\widehat{B}_n \supset \widehat{K}_n \supset \widehat{C}_n \supset \widehat{C}_{n+1}$ y $\widehat{C}_n \downarrow \emptyset$, $n \to +\infty$, $(\widehat{B}_n \downarrow \emptyset$, $n \to +\infty)$.

Además, como P es finitamente aditiva se tiene que, para todo $n \in \mathbb{N}^+$:

$$
\begin{aligned}
P(\widehat{B}_n \setminus \widehat{C}_n) = P(\widehat{B}_n \cap (\widehat{C}_n)^c) = P\left(\widehat{B}_n \cap \left(\bigcup_{i=1}^{n} (\widehat{K}_i)^c\right)\right) = \\
= P\left(\bigcup_{i=1}^{n}\left(\widehat{B}_n \cap (\widehat{K}_i)^c\right)\right) \leqslant \sum_{i=1}^{n} P\left(\widehat{B}_n \cap (\widehat{K}_i)^c\right) \leqslant \\
\leqslant \sum_{i=1}^{n} P\left(\widehat{B}_i \cap (\widehat{K}_i)^c\right) = \sum_{i=1}^{n} P\left(\widehat{B}_i \setminus \widehat{K}i\right) \leqslant \\
\leqslant \sum_{i=1}^{n} \frac{\delta}{2^i} = \delta\left(1 - \frac{1}{2^n}\right) < \delta.
\end{aligned}
$$

Así, puesto que $P(\widehat{B}_n) \geqslant \delta > 0$ y $\widehat{C}_n \subset \widehat{B}_n$, se concluye que $P(\widehat{C}_n) = P(\widehat{B}_n) - P(\widehat{B}_n \setminus \widehat{C}_n) > 0$ y $\widehat{C}_n \neq \emptyset$ para todo $n \in \mathbb{N}^+$.
Sea $\left(x_1^{(n)}, x_2^{(n)}, x_3^{(n)}, ...\right) \in \widehat{C}_n$ (podemos tomar un punto en cada $\widehat{C}_n$, ya que $\widehat{C}_n \neq \emptyset$ para todo $n \in \mathbb{N}^+$). Entonces, $(x_1^n, ..., x_n^n) \in C_n$, $(J_n(C_n) = \widehat{C}_n)$, $n \in \mathbb{N}^+$, (C_n es compacto y cerrado).
Sea (n_1) una subsucesión de (n) tal que $x_1^{(n_1)}$ converge a $x_1^0 \in C_1 = K_1$, lo cual es posible ya que $x_1^n \in C_1$ y C_1 es compacto y cerrado.
Sea (n_2) una subsucesión de (n_1) tal que, $\left(x_1^{(n_2)}, x_2^{(n_2)}\right)$ converge a $\left(x_1^0, x_2^0\right) \in C_2 = (K_1 \times \Omega_2) \cap K_2$, lo cual es posible ya que $\left(x_1^{(n_1)}, x_2^{(n_1)}\right) \in C_2$ y C_2 es compacto y cerrado; el límite precedente recupera x_1^0 (cambiando el límite si es preciso), límite de $x_1^{(n_1)}$ en la etapa anterior, por la propiedad que si una sucesión converge a un punto, entonces toda subsucesión suya converge a ese mismo punto.
Análogamente la sucesión $\left(x_1^{(n_k)}, ..., x_k^{(n_k)}\right)$ converge a $\left(x_1^0, ..., x_k^0\right) \in C_k$, $(C_k = (K_1 \times \Omega_2 \times ... \times \Omega_k) \cap (K_2 \times \Omega_3 \times ... \times \Omega_k) \cap ... \cap K_k)$, compacto y cerrado. Como se ha visto en el segundo paso, en este paso k se reencuentra, (con los cambios que se precisen), $\left(x_1^0, ..., x_{k-1}^0\right)$ límite del paso anterior, de nuevo, por la propiedad que si una sucesión converge a un punto, entonces toda subsucesión suya converge a ese mismo punto.
Así, por inducción se tiene el elemento $(x^0 =)(x_1^0, x_2^0, ...) \in \Omega$ con la propiedad que $x_1^0 \in C_1$, $(x_1^0, x_2^0) \in C_2$,..., $(x_1^0, ..., x_k^0) \in C_k$,..., lo cual implica que $x^0 \in \widehat{C}_n$ para todo $n \in \mathbb{N}^+$, que es una contradicción teniendo en cuenta que $\bigcap_{n \in \mathbb{N}^+} \widehat{C}_n \subset \bigcap_{n \in \mathbb{N}^+} \widehat{B}_n = \emptyset$.
Por tanto, $\lim_{n \to +\infty} P(\widehat{B}_n) = 0$ y se completa la demostración de la σ-aditividad de P sobre $\mathcal{J}$.
El teorema de Carathéodory, (pág. 36 de [4]), concluye la demostración. $\square$

Corolario 4.7. *Sea $\{(X_s, \mathcal{T}_s, \mathcal{F}_s)\}_{s \in N}$, N conjunto infinito numerable, una familia de espacios medibles **s**-polacos, y para cada subconjunto finito F de N sea P_F una*

probabilidad en el espacio medible $\bigotimes_{s\in F}(X_s,\mathcal{B}(X_s))$*, donde* $\mathcal{B}(X_s)=\sigma^{X_s}(\mathcal{T}_s)=\mathcal{F}_s$*,* $(\bigotimes_{s\in F}\mathcal{B}(X_s)=\sigma(\mathcal{A}_F)$*, siendo* $\mathcal{A}_F=\left\{p_s^{-1}(A):\ A\in\mathcal{B}(X_s),\ s\in F\right\})$*. Supongamos que para todo* $F,F'\subset N$*, subconjuntos finitos, con* $F'\subset F$*, se verifica* $P_F\circ p_{FF'}^{-1}=P_{F'}$*, donde* $p_{FF'}:\prod_{s\in F}X_s\to\prod_{s\in F'}X_s$*,* $p_{FF'}(x)=x|_{F'}$*, (es decir,* $p_{FF'}(x)(s)=x(s)$ *para todo* $s\in F'$*),* $x\in\prod_{s\in F}X_s$*. Es claro que* $p_{FF'}$ *es* $\bigotimes_{s\in F}\mathcal{B}(X_s)|\bigotimes_{s\in F'}\mathcal{B}(X_s)$*-medible.*
Entonces, existe una única probabilidad P_N *en*

$$\bigotimes_{s\in N}(X_s,\mathcal{B}(X_s))=\left(\prod_{s\in N}X_s(=\Omega),\bigotimes_{s\in N}\sigma^{X_s}(\mathcal{T}_s)\right)$$

tal que $P_N\circ p_{NF}^{-1}=P_F$ *para todo subconjunto finito* F *de* N*, donde* $p_{NF}:\prod_{s\in N}X_s\to\prod_{s\in F}X_s$*, es la proyección* $p_{NF}(x)=x|_F$*,* $x\in\prod_{s\in N}X_s$*, (la aplicación (proyección)* p_{NF} *es* $\bigotimes_{s\in N}\mathcal{B}(X_s)|\bigotimes_{s\in F}\mathcal{B}(X_s)$*-medible).*
Por último, si N' *es un subconjunto infinito de* N*, entonces* $P_N\circ p_{NN'}^{-1}=P_{N'}$*.*

Teorema 4.8. *Sea* S *un conjunto infinito no numerable. Sean* $(\Omega_s,\mathcal{T}_s)$*,* $s\in S$*, espacios topológicos* **s***-polacos. Se considera* $\mathcal{F}_s=\sigma^{\Omega_s}(\mathcal{T}_s)$ *la* σ*-álgebra generada por* $\mathcal{T}_s$ *en* Ω_s*,* $s\in S$*, y el producto directo de la familia de espacios medibles* $(\Omega_s,\mathcal{F}_s)$*,* $s\in S$*,* $\left(\prod_{s\in S}\Omega_s(=\Omega),\bigotimes_{s\in S}\mathcal{F}_s\right)$*. Sabemos que* $\bigotimes_{s\in S}\mathcal{F}_s=\sigma^{\Omega}(\mathcal{A})$*, donde* $\mathcal{A}=\left\{p_s^{-1}(F):\ F\in\mathcal{F}_s,\ s\in S\right\}$*, y* $\sigma^{\Omega}(\mathcal{A})$ *es la más pequeña* σ*-álgebra en* Ω *que hace medibles las proyecciones* $p_s:\Omega\to\Omega_s$*,* $s\in S$*,* $p_s(\omega)=\omega(s)(=\omega_s)\in\Omega_s$*,* $\omega\in\Omega$*.*
Para cada subconjunto finito F *de* S *se da una probabilidad* P_F *en el espacio medible* $\bigotimes_{s\in F}(\Omega_s,\mathcal{F}_s)=\left(\prod_{s\in F}\Omega_s,\bigotimes_{s\in F}\mathcal{F}_s\right)$*. Supongamos que para todo* $F,F'\subset S$*, subconjuntos finitos, con* $F'\subset F$*, se cumple que* $P_F\circ p_{FF'}^{-1}=P_{F'}$*, donde* $p_{FF'}:\prod_{s\in F}\Omega_s\to\prod_{s\in F'}\Omega_s$*,* $p_{FF'}(x)=x|_{F'}$*, (es decir,* $p_{FF'}(x)(s)=x(s)$ *para todo* $s\in F'$*),* $x\in\prod_{s\in F}\Omega_s$*. Es claro que* $p_{FF'}$ *es* $\bigotimes_{s\in F}\mathcal{F}_s|\bigotimes_{s\in F'}\mathcal{F}_s$*-medible y con la fórmula de consistencia anterior indicamos que, para todo* $B'\in\bigotimes_{s\in F'}\mathcal{F}_s$*,* $P_F\left((p_{FF'})^{-1}(B')\right)=P_{F'}(B')$*.*
Entonces, existe una única probabilidad P *en*

$$\bigotimes_{s\in S}(\Omega_s,\mathcal{F}_s)=\left(\prod_{s\in S}\Omega_s,\bigotimes_{s\in S}\sigma^{\Omega_s}(\mathcal{T}_s)\right)$$

tal que para todo subconjunto finito F*, de* S*, se cumple que* $P\circ p_{SF}^{-1}=P_F$ *, donde* $p_{SF}:\prod_{s\in S}\Omega_s\to\prod_{s\in F}\Omega_s$*,* $p_{SF}(x)=x|_F$*,* $p_{SF}(x)(s)=x(s)$*,* $x\in\prod_{s\in S}\Omega_s$*,* $s\in F$*, (es claro que la proyección* p_{SF} *es* $\bigotimes_{s\in S}\mathcal{F}_s|\bigotimes_{s\in F}\mathcal{F}_s$*-medible).*

Demostración. Sea N un subconjunto de S infinito numerable. Consideramos la subfamilia: $(\Omega_s,\mathcal{T}_s)$, $s\in N$, la σ-álgebra $\mathcal{F}_s=\sigma^{\Omega_s}(\mathcal{T}_s)$, $s\in N$, y el producto directo de los espacios medibles de la subfamilia considerada $\left(\prod_{s\in N}\Omega_s,\bigotimes_{s\in N}\mathcal{F}_s\right)$. Para cada F, subconjunto finito de N (y por tanto subconjunto finito de S) se tiene,

de las hipótesis del teorema presente, una probabilidad P_F en el espacio medible $\left(\prod_{s\in F\subset N}\Omega_s, \bigotimes_{s\in F\subset N}\mathcal{F}_s\right)$ y de estas mismas hipótesis se tiene la propiedad de consistencia:
F y F', subconjuntos finitos de N, con $F'\subset F$, implica que $P_F\circ(p_{FF'})^{-1}=P_{F'}$, donde $p_{FF'}:\prod_{s\in F\subset N}\Omega_s\to\prod_{s\in F'\subset N}\Omega_s$.
A esta subfamilia, obtenida de las hipótesis presentes, le aplicamos el corolario anterior y obtenemos:
Existe una única probabilidad P_N en el espacio medible $\left(\prod_{s\in N}\Omega_s, \bigotimes_{s\in N}\mathcal{F}_s\right)$ tal que para todo subconjunto finito F de N se verifica que $P_F = P_N\circ(p_{NF})^{-1}$, $(p_{NF}:\prod_{s\in N}\Omega_s\to\prod_{s\in F}\Omega_s,\ p_{NF}(x)=x|_F,\ x\in\prod_{s\in N}X_s)$. Observamos que si N_1 y N_2 son subconjuntos infinitos numerables de S con $N_1\subset N_2$, entonces $P_{N_1}=P_{N_2}\circ(p_{N_2N_1})^{-1}$ (véase el corolario anterior).
Luego de lo visto hasta aquí, se tiene el siguiente aserto: *Para todo subconjunto $N\subset S$, infinito numerable, existe una única probabilidad P_N en $\left(\prod_{s\in N}\Omega_s, \bigotimes_{s\in N}\mathcal{F}_s\right)$ tal que para todo $F\subset N$, finito, $P_F=P_N\circ p_{NF}^{-1}$. Además, si $N_1\subset N_2$, N_1 y N_2 subconjuntos infinito numerables de S, entonces $P_{N_1}=P_{N_2}\circ p_{N_2N_1}^{-1}$.*
Para todo $B\in\bigotimes_{s\in S}\mathcal{F}_s$, por el **Corolario 3.1.54.**, (pág. 52), de [4], existe un subconjunto numerable infinito N de S y existe $B_N\in\bigotimes_{s\in N}\mathcal{F}_s$ tal que $B=(p_{SN})^{-1}(B_N)$. Definimos $P(B)=P_N(B_N)$. Veamos que esta definición no depende de la representación anterior de B.
Sea N' otro subconjunto numerable infinito de S y $B_{N'}\in\bigotimes_{s\in N'}\mathcal{F}_s$ tales que $B=(p_{SN'})^{-1}(B_{N'})$, en este caso tendríamos $P(B)=P_{N'}(B_{N'})$. Sea $N''=N\cup N'$, (subconjunto infinito numerable de S). Por la observación anterior tenemos que $P_N = P_{N''}\circ(p_{N''N})^{-1}$ y $P_{N'}=P_{N''}\circ(p_{N''N'})^{-1}$. Por otro lado,

$$
\begin{aligned}
B=(p_{SN})^{-1}(B_N)=(p_{SN''})^{-1}(p_{N''N})^{-1}(B_N)=&\\
&=(p_{SN''})^{-1}(p_{N''N'})^{-1}(B_{N'})=(p_{SN'})^{-1}(B_{N'}),
\end{aligned}
$$

lo cual implica que $(p_{N''N})^{-1}(B_N)=(p_{N''N'})^{-1}(B_{N'})$, ($p_{SN''}$ es suprayectiva). Así,

$$P(B)=P_N(B_N)=P_{N''}(p_{N''N})^{-1}(B_N)=P_{N''}(p_{N''N'})^{-1}(B_{N'})=P_{N'}(B_{N'}).$$

Se tiene por tanto definida una aplicación P sobre $\bigotimes_{s\in S}\mathcal{F}_s$ cumpliendo que $P\circ p_{SF}^{-1}= P_F$ para todo subconjunto finito F de S. En efecto: Sea $F\subset S$, finito, y N infinito numerable tal que $F\subset N\subset S$, y sea $B_N\in\bigotimes_{s\in N}\mathcal{F}_s$.

$$
\begin{aligned}
&\text{Entonces,}\ \ p_{SN}^{-1}(B_N)(=B)\in\bigotimes_{s\in S}\mathcal{F}_s,\ \ P(B)=P_N(B_N),\ \ P_F=P_N\circ p_{NF}^{-1},\\
&\qquad P(B)=P(p_{SN}^{-1}(B_N))=P_N(B_N)\ \text{ y por tanto }\ P\circ p_{SN}^{-1}=P_N.\ \text{ Así,}\\
&\qquad\qquad\qquad P\circ p_{SN}^{-1}\circ p_{NF}^{-1}=P_N\circ p_{NF}^{-1}\ \text{ y }\ P\circ p_{SF}^{-1}=P_F.
\end{aligned}
$$

Se tiene que P es una probabilidad en $\left(\prod_{s\in S}\Omega_s, \bigotimes_{s\in S}\mathcal{F}_s\right)$. En efecto:
(1). Si N es un subconjunto infinito numerable del conjunto S, entonces $P(\prod_{s\in S}\Omega_s)= P_N(\prod_{s\in N}\Omega_s)=1$.

(2). Sean B_1, B_2,..., elementos disjuntos dos a dos de $\bigotimes_{s\in S}\mathcal{F}_s$, y para cada $i = 1, 2, ...$ sean N_i subconjunto infinito numerable de S y $B_{N_i} \in \bigotimes_{s\in N_i}\mathcal{F}_s$ tales que $B_i = (p_{SN_i})^{-1}(B_{N_i})$, (**Corolario 3.1.54.**, pág. 52, de [4]). Entonces, $N = \bigcup_{i=1}^{\infty} N_i$ es un subconjunto infinito numerable de S y para todo $i = 1, 2, ..., B'_i = (p_{NN_i})^{-1}(B_{N_i})$ es un elemento de $\bigotimes_{s\in N}\mathcal{F}_s$ tal que $B_i = (p_{SN})^{-1}(B'_i)$. Es claro que B'_1, B'_2,..., son disjuntos dos a dos y $(p_{SN})^{-1}(B'_1 \cup B'_2 \cup ...) = B_1 \cup B_2 \cup ....$ Así,

$$P\left(\bigcup_{i=1}^{\infty} B_i\right) = P_N\left(\bigcup_{i=1}^{\infty} B'_i\right) = \sum_{i=1}^{\infty} P_N(B'_i) = \sum_{i=1}^{\infty} P(B_i).$$

Finalmente P es única tal que $P \circ p_{SF}^{-1} = P_F$ para todo subconjunto finito F de S. En efecto: Sea P' otra probabilidad en $\bigotimes_{s\in S}(\Omega_s, \mathcal{F}_s)$ con $P' \circ p_{SF}^{-1} = P_F$ para toda parte finita F de S. Entonces, P y P' coinciden sobre el álgebra,

$$\mathcal{A}^* = \left\{p_{SF}^{-1}(B) : F \subset S, \text{ finito}, \ B \in \bigotimes_{s\in F}\mathcal{F}_s\right\} \text{ sobre } \Omega.$$

Además, $\mathcal{A} \subset \mathcal{A}^*$. Finalmente, como $\sigma^{\Omega}(\mathcal{A}) = \bigotimes_{s\in S}\mathcal{F}_s$, por el teorema de Carathéodory (pág. 36 de [4]), $P = P'$. □

5. Convergencia débil y convergencia en distribución

Sea $(X, \mathcal{T})$ un espacio topológico. Se designa por $C_a(X)$ al conjunto de funciones continuas y acotadas de $(X, \mathcal{T})$ en $(\mathbb{R}, \mathcal{T}_u)$.
Se verifica que toda $f \in C_a(X)$ es integrable respecto a cualquier medida de Borel finita en $(X, \mathcal{B}(X))$.
Dada una sucesión $\{\mu_n\}_{n\in\mathbb{N}}$ de medidas de Borel finitas en $(X, \mathcal{B}(X))$, se dice que $\{\mu_n\}_{n\in\mathbb{N}^+}$ *converge débilmente a* μ_0, y se escribirá $\mu_n \Rightarrow \mu_0$, si $\lim_n \int_X f d\mu_n = \int_X f d\mu_0$ para toda $f \in C_a(X)$.

Sean $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y $\{\xi_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias de $(\Omega, \mathcal{F}, P)$ en $(X, \mathcal{B}(X))$. Se dice que $\{\xi_n\}_{n\in\mathbb{N}^+}$ *converge en distribución a* ξ_0, y se escribe $\xi_n \Rightarrow \xi_0$, si la sucesión de probabilidades $\{P_{\xi_n}\}_{n\in\mathbb{N}^+}$ converge débilmente a la probabilidad P_{ξ_0}, $(P_{\xi_n}(B) = P((\xi_n)^{-1}(B))$, $n \in \mathbb{N}$, $B \in \mathcal{B}(X))$.

Proposición 5.1 (**Unicidad del límite**). *Sean $(X, \mathcal{T})$ un espacio topológico perfectamente normal y $\{P_n\}_{n\in\mathbb{N}^+}$ una sucesión de probabilidades que converge débilmente a la probabilidad P y a la probabilidad Q. Entonces, $P = Q$. Por tanto, si $\{\xi_n\}_{n\in\mathbb{N}^+}$ es una sucesión de variables aleatorias de un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ en $(X, \mathcal{B}(X))$ que converge en distribución a ξ_0 y a ξ'_0, se tiene que estas variables aleatorias tienen la misma distribución de probabilidad en $(X, \mathcal{B}(X))$, $(P_{\xi_0} = P_{\xi'_0})$.*

Demostración. Sea C un cerrado no vacío de $(X,\mathcal{T})$. Como $(X,\mathcal{T})$ es perfectamente normal, C es G_δ y por tanto existe $\{A_n : n \in \mathbb{N}^+\} \subset \mathcal{T}$ tal que $C = \bigcap_n A_n$ y $A_{n+1} \subset A_n$, $n \in \mathbb{N}^+$. Ahora, para todo $n \in \mathbb{N}^+$, por el **Corolario 4.3.50.**, pág. 232, de [6], existe una función continua f_n de X en $[0,1]$ tal que $f_n^{-1}(1) = C$ y $f_n^{-1}(0) = A_n^c$. Es claro que $\lim_n f_n = I_C$. Entonces, por el teorema de la convergencia dominada (**Teorema 3.4.10.**, pág. 172, de [4]), se tiene que: $\lim_n \lim_m \int_X f_n dP_m = \lim_n \int_X f_n dP = \int_X I_C dP = P(C)$ y $\lim_n \lim_m \int_X f_n dP_m = \lim_n \int_X f_n dQ = \int_X I_C dQ = Q(C)$. Así, $P(C) = Q(C)$ para todo $C \in \mathcal{C}_\mathcal{T}$, y, por la **Proposición 4.3**, $P = Q$. □

Teorema 5.2 (**Portmanteau**). *Sean $(X,\mathcal{T})$ un espacio topológico perfectamente normal, (en particular, un espacio topológico seudometrizable), y $\{P_n\}_{n\in\mathbb{N}}$ una sucesión de probabilidades en $(X,\mathcal{B}(X))$, ($\mathcal{B}(X) = \sigma^X(\mathcal{T})$). Las siguientes afirmaciones son equivalentes:*

(1). $P_n \Rightarrow P_0$.

(2). $\limsup_n P_n(C) \leqslant P_0(C)$, *para todo cerrado C de $(X,\mathcal{T}_d)$.*

(3). $\liminf_n P_n(G) \geqslant P_0(G)$, *para todo abierto G de $(X,\mathcal{T}_d)$.*

(4). $\lim_n P_n(B) = P_0(B)$, *para todo $B \in \mathcal{B}(X)$ con $P_0(Fr(B)) = 0$.*

Demostración. **(1)**$\Longrightarrow$ **(2)**. Si C es vacío, el resultado de **(2)** es trivial. Sea C un cerrado no vacío de $(X,\mathcal{T})$. Como $(X,\mathcal{T})$ es perfectamente normal, C es G_δ y por tanto existe una sucesión de abiertos $\{G_n\}_{n\in\mathbb{N}^+}$ con $C = \bigcap_{n\in\mathbb{N}^+} G_n$ y $G_{n+1} \subset G_n$ para todo $n \in \mathbb{N}^+$. De nuevo, como $(X,\mathcal{T})$ es perfectamente normal, para todo $n \in \mathbb{N}^+$, por el **Corolario 4.3.50.**, pág. 232, de [6], existe una función continua f_n de $(X,\mathcal{T})$ en $[0,1]$, con su topología usual, tal que $f_n^{-1}(1) = C$ y $f_n^{-1}(0) = G_n^c$.
Puesto que $P_n(C) = \int_X I_C dP_n \leqslant \int_X f_m dP_n$, ($I_C \leqslant f_m$), obtenemos, por **(1)**, que para todo $m \in \mathbb{N}^+$,

$$\limsup_n P_n(C) \leqslant \limsup_n \int_X f_m dP_n = \int_X f_m dP_0 \leqslant \int_X I_{G_m} dP_0 = P_0(G_m).$$

Ahora por el **Teorema 3.1.11.**, pág. 12, de [4], se tiene que $\lim_m P_0(G_m) = P_0(C)$, y por tanto, $\limsup_n P_n(C) \leqslant \lim_m P_0(G_m) = P_0(C)$.
(2)$\Longleftrightarrow$**(3)**. Se pasa de un resultado al otro tomando complementarios.
(3)$\Longrightarrow$**(4)**. Sea $B \in \mathcal{B}(X)$ con $P_0(Fr(B)) = 0$. Tenemos que $Int(B) \subset B \subset \overline{B}$. Así, por **(3)** ($\Longleftrightarrow$ **(2)**), se tiene:

$$\begin{aligned} \limsup_n P_n(B) &\leqslant \limsup_n P_n(\overline{B}) \leqslant P_0(\overline{B}) = P_0(B \cup Fr(B)) = P_0(B), \\ \liminf_n P_n(B) &\geqslant \liminf_n P_n(Int(B)) \geqslant P_0(Int(B)) = P_0(B \setminus Fr(B)) = P_0(B). \end{aligned}$$

Por tanto, $\lim_n P_n(B) = P_0(B)$.
(4)$\Longrightarrow$**(1)**. Sea $g \in C_a(X)$. Entonces, g es una variable aleatoria de $(X,\mathcal{B}(X))$ en

$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, y por tanto, se tiene la probabilidad π_g en $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ definida por $\pi_g(A) = P_0(g^{-1}(A))$, $A \in \mathcal{B}(\mathbb{R})$.
Como g está acotada, existen $a, b \in \mathbb{R}$ con $a < b$ e $im(g) \subset (a, b)$, y por consiguiente $\pi_g(\mathbb{R} \setminus (a, b)) = 0$. Por otro lado, π_g es una probabilidad y así por el **Lema 4.1**, existe a lo más una cantidad numerable de números reales r tales que $\pi_g(\{r\}) \neq 0$. Luego, para todo $\varepsilon > 0$ existen números reales t_0, t_1,..., t_m tales que: $a = t_0 < t_1 < ... < t_m = b$; $t_{i+1} - t_i < \varepsilon$, $i = 0, 1, ..., m-1$; y $\pi_g(\{t_i\}) = 0$, $i = 0, 1, ..., m$.
Para todo $i \in \{0, 1, ..., m-1\}$, sea $B_i = g^{-1}\left([t_i, t_{i+1})\right) \in \mathcal{B}(X)$. Es claro que $X = \bigcup_{i=0}^{m-1} B_i$ y $P_0(Fr(B_i)) = 0$, $(g^{-1}((t_i, t_{i+1})) \subset Int(B_i) \subset B_i \subset \overline{B_i} \subset g^{-1}([t_i, t_{i+1}])$ implica que $P_0(Fr(B_i)) = P_0(\overline{B_i} \setminus Int(B_i)) \leqslant P_0(\{x : x \in X,\ \text{y } g(x) = t_i \text{ o } g(x) = t_{i+1}\}) = \pi_g(\{t_i\}) + \pi_g(\{t_{i+1}\}) = 0)$, $i = 0, 1, ..., m-1$. Por tanto, por la hipótesis **(4)**, $\lim_n P_n(B_i) = P_0(B_i)$ para todo $i \in \{0, 1, ..., m-1\}$.
Consideramos la variable aleatoria simple $h = \sum_{i=0}^{m-1} t_i I_{B_i} : X \to \mathbb{R}$, la cual cumple que $h(x) \leqslant g(x) \leqslant h(x) + \varepsilon$ para todo $x \in X$. Entonces,

$$\left|\int_X g dP_n - \int_X g dP_0\right| \leqslant \int_X |g-h| dP_n + \left|\int_X h dP_n - \int_X h dP_0\right| + \int_X |g-h| dP_0 \leqslant$$
$$\leqslant 2\varepsilon + \left|\sum_{i=0}^{m-1} t_i \left(P_n\left(B_i\right) - P_0\left(B_i\right)\right)\right|,$$

lo cual prueba que $\lim_n \int_X g dP_n = \int_X g dP_0$. □

Teniendo en cuenta el teorema del cambio de variable en integrales (**Teorema 3.4.28.**, pág. 189, de [4]), del teorema anterior se deduce

Teorema 5.3 (**Portmanteau**). *Sean $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, $(X, \mathcal{T})$ un espacio topológico perfectamente normal, (en particular, un espacio topológico seudo-metrizable), y $\{\xi_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias de $(\Omega, \mathcal{F}, P)$ en $(X, \mathcal{B}(X))$. Las siguientes afirmaciones son equivalentes:*

(1). $\xi_n \Rightarrow \xi_0$.

(2). $\limsup_n P(\xi_n \in C) \leqslant P(\xi_0 \in C)$, *para todo cerrado C de $(X, \mathcal{T}_d)$.*

(3). $\liminf_n P(\xi_n \in G) \geqslant P(\xi_0 \in G)$, *para todo abierto G de $(X, \mathcal{T}_d)$.*

(4). $\lim_n P(\xi_n \in B) = P(\xi_0 \in B)$, *para todo $B \in \mathcal{B}(X)$ con $P(\xi_0 \in Fr(B)) = 0$.*

A continuación se establece que la convergencia débil de probabilidades, y como consecuencia la convergencia en distribución de variables aleatorias, se puede describir (con algunas restricciones) como convergencia en un espacio métrico.
Sea (X, d) un espacio seudométrico. Se designa por $\mathbf{P}(X)$ al conjunto de probabilidades en $(X, \mathcal{B}(X))$ y se define $d_P : \mathbf{P}(X) \times \mathbf{P}(X) \to [0, +\infty)$ por:

$$d_P(P, Q) = \inf\{\varepsilon > 0 : P(A) \leqslant Q(B(A; \varepsilon)) + \varepsilon \text{ y } Q(A) \leqslant P(B(A; \varepsilon)) + \varepsilon,\ A \in \mathcal{B}(X)\},$$

donde $B(A; \varepsilon) = \{x \in X : d(x, A) < \varepsilon\}$ si $A \neq \emptyset$ y $B(\emptyset; \varepsilon) = \emptyset$, $P, Q \in \mathbf{P}(X)$.

Proposición 5.4. *Sean (X,d) un espacio seudométrico. Se considera la relación de equivalencia $\sim$ en X dada por: $x \sim y \iff d(x,y) = 0$. En el conjunto cociente $X^* = X/\sim$ se tiene la métrica d^* definida por: $d^*([x],[y]) = d(x,y)$, $x,y \in X$. Sea $p : X \to X^*$ la proyección natural, $p(x) = [x]$, $x \in X$, la cual es una identificación, (p es suprayectiva y $\mathcal{T}_{d^*} = \{G^* : G^* \subset X^* \text{ y } p^{-1}(G^*) \in \mathcal{T}_d\}$), abierta y cerrada de $(X,\mathcal{T}_d)$ en $(X^*,\mathcal{T}_{d^*})$, ($p^{-1}(p(G)) = G$ para todo $G \in \mathcal{T}_d$). Se verifica:*

(a). *Sean $\mathbf{P}(X)$ y $\mathbf{P}(X^*)$ el conjunto de probabilidades en $(X,\mathcal{B}(X))$ y $(X^*,\mathcal{B}(X^*))$, respectivamente. Entonces, la aplicación $\varphi : \mathbf{P}(X) \to \mathbf{P}(X^*)$, definida por $\varphi(P) = P_p$, donde $P_p(B^*) = P(p^{-1}(B^*))$, $P \in \mathbf{P}(X)$, $B^* \in \mathcal{B}(X^*)$, (p es variable aleatoria de $(X,\mathcal{B}(X))$ en $(X^*,\mathcal{B}(X^*))$), es biyectiva.*

(b). *Para todo $P,Q \in \mathbf{P}(X)$, se tiene que $d_P(P,Q) = d_P^*(\varphi(P),\varphi(Q))$.*

(c). *Para todo $x \in X$, $\varphi(\delta_x) = \delta_{[x]}$, y por tanto, para todo $x,y \in X$, $\delta_x = \delta_y$ si y sólo si $d(x,y) = 0$. Además, $d_P(\delta_x,\delta_y) = d_P^*(\delta_{[x]},\delta_{[y]})$ para todo $x,y \in X$.*

Demostración. **(a).** Sean $P,Q \in \mathbf{P}(X)$ tales que $\varphi(P) = \varphi(Q)$. Por la **Proposición 2.3(3)**, y teniendo en cuenta que $\mathcal{T}_d = p^{-1}(\mathcal{T}_{d^*})$, ($p$ es una identificación, y $p^{-1}(p(G)) = G$, $G \in \mathcal{T}_d$), se verifica que

$$\mathcal{B}(X) = \sigma^X(\mathcal{T}_d) = \sigma^X(p^{-1}(\mathcal{T}_{d^*})) = p^{-1}(\sigma^{X^*}(\mathcal{T}_{d^*})) = p^{-1}(\mathcal{B}(X^*)).$$

Así, para todo $B \in \mathcal{B}(X)$, existe $B^* \in \mathcal{B}(X^*)$ tal que $B = p^{-1}(B^*)$, y por tanto, $P(B) = P(p^{-1}(B^*)) = \varphi(P)(B^*) = \varphi(Q)(B^*) = Q(p^{-1}(B^*)) = Q(B)$. Luego, $P = Q$ y φ es inyectiva.
Veamos ahora que φ es suprayectiva. Sea $P^* \in \mathcal{B}(X^*)$. Se define $P : \mathcal{B}(X) \to \mathbb{R}$ por $P(B) = P^*(B^*)$, donde $B^* \in \mathcal{B}(X^*)$ y cumple que $p^{-1}(B^*) = B$, ($\mathcal{B}(X) = p^{-1}(\mathcal{B}(X^*))$). Como p es suprayectiva, dado $B \in \mathcal{B}(X)$, existe un único $B^* \in \mathcal{B}(X^*)$ con $B = p^{-1}(B^*)$.
Se tiene que P es una probabilidad en $(X,\mathcal{B}(X))$. En efecto: Es claro que $P(\emptyset) = P^*(\emptyset) = 0$, $P(B) = P^*(B^*) \geqslant 0$, $B \in \mathcal{B}(X)$, $P(X) = P^*(X^*) = 1$.
Finalmente, sea $\{B_n\}_{n\in\mathbb{N}} \subset \mathcal{B}(X)$ con $B_n \cap B_m$ para todo $n,m \in \mathbb{N}$ con $n \neq m$. Para cada $n \in \mathbb{N}$ sea $B_n^* \in \mathcal{B}(X^*)$ tal que $B_n = p^{-1}(B_n^*)$. Se tiene que para todo $n,m \in \mathbb{N}$ con $n \neq m$, $B_n^* \cap B_m^* = \emptyset$, ya que si $z \in B_n^* \cap B_m^*$ existe $x \in B_n$ y existe $y \in B_m$ tales que $p(x) = p(y) = z$, lo cual implica que $d(x,y) = 0$. Como $d(x,y) = 0$, es claro que para todo $G \in \mathcal{T}_d$, $\{x,y\} \subset G$ o $\{x,y\} \subset G^c$, y puesto que $\sigma^X(\mathcal{T}_d) = \mathcal{B}(X)$, por el **Lema 2.2**, $\{x,y\} \subset B$ o $\{x,y\} \subset B^c$ para todo $B \in \mathcal{B}(X)$. Así, $\{x,y\} \subset B_n \cap B_m$, lo cual es absurdo. Entonces, teniendo en cuenta que $p^{-1}\left(\bigcup_{n\in\mathbb{N}} B_n^*\right) = \bigcup_{n\in\mathbb{N}} B_n$,

$$P\left(\bigcup_{\nu\in\mathbb{N}} B_\nu\right) = P^*\left(\bigcup_{\nu\in\mathbb{N}} B_\nu^*\right) = \sum_{\nu\in\mathbb{N}} P^*\left(B_\nu^*\right) = \sum_{\nu\in\mathbb{N}} P\left(B_\nu\right).$$

Así, P es una probabilidad en el espacio medible $(X,\mathcal{B}(X))$ y se comprueba fácilmente que $\varphi(P) = P^*$.

(b). Basta observar que para todo $A^* \in \mathbb{B}(X^*)$ y todo $\varepsilon > 0$, $p^{-1}(B(A^*;\varepsilon)) = B\left(p^{-1}(A^*);\varepsilon\right)$ y que $p^{-1}(\mathcal{B}(X^*)) = \mathcal{B}(X)$.

(c). Se comprueba fácilmente a partir de la definición de las probabilidades de Dirac y de los resultados **(a)** y **(b)**. $\square$

Teorema 5.5 (**Prokhorov**). *Sea (X,d) un espacio seudométrico.*

(1). *d_P es una métrica en $\mathbf{P}(X)$ tal que $d_P(\delta_x,\delta_y) = \text{mín}\{d(x,y),1\}$ para todo $x,y \in X$.*

(2). *Si $\{P_n\}_{n\in\mathbb{N}^+}$ converge a $P_0 \in \mathbf{P}(X)$ en $(\mathbf{P}(X),d_P)$, entonces, $P_n \Rightarrow P_0$.*

(3). *Si (X,d) es separable, entonces $(\mathbf{P}(X),d_P)$ es separable. Además, en este caso, si $P_n \Rightarrow P_0$, la sucesión de probabilidades $\{P_n\}_{n\in\mathbb{N}^+}$ converge a P_0 en $(\mathbf{P}(X),d_P)$.*

(4). *Si (X,d) es separable: (X,d) es completo si y sólo si $(\mathbf{P}(X),d_P)$ es completo.*

(5). *Supongamos que (X,d) es separable y completo, y sea $\Pi \subset \mathbf{P}(X)$. Entonces, $\overline{\Pi}$ es compacto en $(\mathbf{P}(X),d_P)$ si y sólo si para todo $\varepsilon > 0$ existe un subconjunto cerrado y compacto, K_ε, en $(X,\mathcal{T}_d)$ tal que $P(K_\varepsilon) \geqslant 1-\varepsilon$ para todo $P \in \Pi$.*

Demostración. La prueba se obtiene de la proposición anterior y del teorema de Prokhorov, (véase [1]), aplicado al espacio métrico (X^*,d^*) considerado en la citada proposición. $\square$

6. Procesos estocásticos

Sean $(\Omega,\mathcal{F})$, $(E,\mathcal{E})$ espacios medibles y S un conjunto infinito. Una familia de variables aleatorias $\xi_s : \Omega \to E$, $s \in S$, de $(\Omega,\mathcal{F})$ en $(E,\mathcal{E})$, se llama *proceso estocástico en* $(\Omega,\mathcal{F})$ *con dominio del parámetro* S y *espacio de estados* $(E,\mathcal{E})$. En este caso escribiremos $\tilde{\xi} = \{\xi_s\}_{s\in S}$.

Sea $\tilde{\xi} = \{\xi_s\}_{s\in S}$ un proceso estocástico en $(\Omega,\mathcal{F})$, con dominio de parámetro S y espacio de estados $(E,\mathcal{E})$. Para cada $\omega \in \Omega$, a la función

$$\begin{array}{rcl} \varphi_{\tilde{\xi}}^{\omega} : S & \to & E \\ s & \mapsto & \xi_s(\omega) \end{array}$$

se le llama *trayectoria o realización* del proceso estocástico correspondiente a ω. Se tiene la aplicación $\varphi_{\tilde{\xi}} : \Omega \to E^S$ definida por $\varphi_{\tilde{\xi}}(\omega) = \varphi_{\tilde{\xi}}^{\omega}$, $\omega \in \Omega$, la cual cumple que $p_s \circ \varphi_{\tilde{\xi}} = \xi_s$ para cada $s \in S$, (p_s es la proyección $p_s(x) = x(s) = x_s$, $s \in S$, $x \in E^S$). Si consideramos el espacio medible producto directo de la familia de espacios medibles $(E_s,\mathcal{E}_s) = (E,\mathcal{E})$, $s \in S$, $(E^S, \bigotimes_{s\in S}\mathcal{E}_s)$, entonces $\varphi_{\tilde{\xi}}$ es una variable aleatoria de $(\Omega,\mathcal{F})$ en $(E^S, \bigotimes_{s\in S}\mathcal{E}_s)$, (consecuencia de la **Proposición 2.6**).

Recíprocamente, si η es una variable aleatoria de $(\Omega, \mathcal{F})$ en $(E^S, \bigotimes_{s\in S}\mathcal{E}_s)$, entonces $\tilde{\eta} = \{\eta_s = p_s \circ \eta\}_{s\in S}$ es un proceso estocástico en $(\Omega, \mathcal{F})$ con dominio del parámetro S y espacio de estados $(E, \mathcal{E})$.

Si $(E, \mathcal{E})$ es el espacio de Borel de un espacio topológico $(X, \mathcal{T})$, (es decir, $E = X$ y $\mathcal{E} = \sigma^X(\mathcal{T})$), y las trayectorias de $\tilde{\xi}$ cumplen alguna propiedad de regularidad (continuidad, variación acotada, etc.), es importante analizar si se puede reducir el espacio $(E^S, \bigotimes_{s\in S}\mathcal{E}_s)$ a uno que sea de Borel de un espacio topológico y contenga la imagen de $\varphi_{\tilde{\xi}}$. A continuación se analiza el caso de trayectorias continuas en el que la contestación es afirmativa.

Procesos estocásticos con trayectorias continuas

Teorema 6.1. *Sean $(S, \mathcal{T}')$ un espacio topológico compacto de Hausdorff infinito y II.A.N., $(X, \mathcal{T})$ un espacio topológico **s**-polaco y $C(S, X)$ el conjunto de las aplicaciones continuas de $(S, \mathcal{T}')$ en $(X, \mathcal{T})$, $(C(S, X) \subset X^S)$. Entonces,*

(1). *$(C(S, X), \mathcal{T}_c)$, donde $\mathcal{T}_c$ es la topología compacta-abierta en $C(S, X)$, es un espacio topológico **s**-polaco.*

(2). *La σ-álgebra en $C(S, X)$, $\left(\bigotimes_{s\in S}(\sigma^X(\mathcal{T}))_s\right) \bigcap C(S, X) (= \mathcal{E})$, ((2*) de la **Proposición 2.3**), donde $(\sigma^X(\mathcal{T}))_s = \sigma^X(\mathcal{T})$, $s \in S$, coincide con la σ-álgebra $\sigma^{C(S,X)}(\mathcal{T}_c) = \mathcal{B}(C(S, X))$.*

Demostración. **(1).** Sea d una seudométrica completa y separable en X tal que $\mathcal{T}_d = \mathcal{T}$. En $C(S, X)$ consideramos la seudométrica $\rho_d(\varphi, \psi) = \sup_{s\in S} d(\varphi(s), \psi(s))$, (como los elementos de $C(S, X)$ son funciones continuas y $(S, \mathcal{T}')$ es compacto, $\rho_d(\varphi, \psi) = \text{máx}_{s\in S} d(\varphi(s), \psi(s))$), y sea $\mathcal{T}_{\rho_d}$ la topología determinada por esta seudométrica ρ_d. Entonces, por la **Proposición 7.1.23.(5)**, pág. 360 de [6], $(C(S, X), \rho_d)$ es un espacio seudométrico completo y $\mathcal{T}_{\rho_d} = \mathcal{T}_c$, y por el **Corolario 7.1.10.**, pág. 350 de [6], $\mathcal{T}_{\rho_d} = \mathcal{T}_c$ tiene una base numerable. Así, $(C(S, X), \mathcal{T}_{\rho_d}) = (C(S, X), \mathcal{T}_c)$ es un espacio topológico **s**-polaco.

(2). (a). Para todo $x \in X$, todo $\varepsilon > 0$, y todo $s \in S$, $\{\varphi \in C(S, X) : \varphi(s) \in B(x; \varepsilon)\} \in \mathcal{T}_c$. Así, se tiene la inclusión $\mathcal{E} \subset \sigma^{C(S,X)}(\mathcal{T}_c)$, pues por la **Proposición 2.5**, $\sigma^{X^S}\left(\{\psi \in X^S : \psi(s) \in B(x; \varepsilon)\} : x \in X, s \in S, \varepsilon > 0\right\} = \bigotimes_{s\in S}\left(\sigma^X(\mathcal{T})\right)_s$, (obsérvese que al ser $(X, \mathcal{T})$ **s**-polaco, se verifica que $\sigma^X(\{B(x; \varepsilon) : x \in X, \varepsilon > 0\}) = \sigma^X(\mathcal{T})$).
(b). Como $(S, \mathcal{T}')$ cumple el II.A.N., existe un subconjunto numerable $D = \{s_n\}_{n\in\mathbb{N}}$ de S denso en $(S, \mathcal{T}')$. Entonces, para todo $\varepsilon > 0$ y todo $\varphi_0 \in C$,

$$B^{=}(\varphi_0; \varepsilon) = \{\varphi \in C(S, X) : \rho_d(\varphi, \varphi_0) \leqslant \varepsilon\} = \\ = \bigcap_{n\in\mathbb{N}} \{\varphi \in C(S, X) : d(\varphi(s_n), \varphi_0(s_n)) \leqslant \varepsilon\} \in \mathcal{E},$$

Por tanto, $\sigma^{C(S,X)}(\mathcal{T}_{\rho_d}) \subset \mathcal{E}$, ya que en espacios topológicos **s**-polacos, la σ-álgebra generada por las bolas cerradas es igual a la σ-álgebra generada por las bolas abiertas, (véase la descripción de σ-álgebras mediante seudométricas). $\square$

Proposición 6.2. *Sean $(\Omega, \mathcal{F})$ un espacio medible, $C(S,X)$ el conjunto de aplicaciones continuas de un espacio compacto de Hausdorff infinito y II.A.N., $(S, \mathcal{T}')$, en un espacio topológico **s**-polaco $(X, \mathcal{T})$, y $(C(S,X), \mathcal{B}(C(S,X)))$ el espacio medible construido en el teorema anterior. Entonces:*

(1) *Sea $\xi : \Omega \to C(S,X)$ una variable aleatoria de $(\Omega, \mathcal{F})$ en el espacio medible $(C(S,X), \mathcal{B}(C(S,X)))$. Ponemos $\xi_s = p_s \circ \xi$, $s \in S$, donde $p_s : C(S,X) \to X$ es la proyección $p_s(\varphi) = \varphi_s = \varphi(s)$, $\varphi \in C(S,X)$. Entonces, $\xi_s : \Omega \to X$ es variable aleatoria de $(\Omega, \mathcal{F})$ en $(X, \mathcal{B}(X))$, $s \in S$, y para todo $\omega \in \Omega$, la aplicación $\varphi^\omega : S \to X$, $\varphi^\omega(s) = \xi_s(\omega)$, $s \in S$, es continua (**Proposición 7.1.31.**, pág. 364 de [6]). Por tanto, $\tilde{\xi} = \{\xi_s\}_{s \in S}$ es un proceso estocástico en $(\Omega, \mathcal{F})$ con dominio del parámetros S y espacio de estados $(X, \mathcal{B}(X))$ con trayectorias continuas, (véase la **Proposición 2.6**).*

(2) *Sea $\tilde{\eta} = \{\eta_s : \Omega \to X\}_{s \in S}$, un proceso estocástico en $(\Omega, \mathcal{F})$ con dominio del parámetro S y espacio de estados $(X, \mathcal{B}(X))$, y con trayectorias continuas ($s \mapsto \eta_s(\omega)$ es continua para cada ω). Entonces, la aplicación $\psi_{\tilde{\eta}} : \Omega \to C(S,X)$, definida por $(\psi_{\tilde{\eta}}(\omega))(s) = \eta_s(\omega)$, $\omega \in \Omega$, $s \in S$, es una variable aleatoria de $(\Omega, \mathcal{F})$ en $(C(S,X), \mathcal{B}(C(S,X)))$, (obsérvese que $p_s \circ \psi_{\tilde{\eta}} = \eta_s$, $s \in S$).*
*(Para la demostración de este apartado **(2)**, basta tener en cuenta que $j \circ \psi_{\tilde{\eta}} = \varphi_{\tilde{\eta}} : \Omega \to X^S$, $j : C(S,X) \hookrightarrow X^S$, es variable aleatoria de $(\Omega, \mathcal{F})$ en $(X^S, \bigotimes_{s \in S} \left(\sigma^X(\mathcal{T})\right)_s)$, y que $\mathcal{E} = \sigma^{C(S,X)}(\mathcal{T}_c)$.)*

(3) *Dos procesos estocásticos $\tilde{\eta}^1 = \left\{\eta_s^1 : \Omega \to X\right\}_{s \in S}$ y $\tilde{\eta}^2 = \left\{\eta_s^2 : \Omega \to X\right\}_{s \in S}$ en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ con dominio del parámetro S y espacio de estados $(X, \mathcal{B}(X))$, y con trayectorias continuas, tienen la misma ley de probabilidad (véase el párrafo que sigue a esta proposición) si y sólo si $\psi_{\tilde{\eta}^1}$ y $\psi_{\tilde{\eta}^2}$ inducen la misma probabilidad en $(C(S,X), \mathcal{B}(C(S,X)))$.*

Sean $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, $(E, \mathcal{E})$ un espacio medible y S un conjunto infinito. Sea $\tilde{\xi} = \{\xi_s\}_{s \in S}$ un proceso estocástico en $(\Omega, \mathcal{F}, P)$ con dominio del parámetro S y espacio de estados $(E, \mathcal{E})$. Para cada $F \subset S$, finito y no vacío, se tiene la probabilidad $P_F^{\tilde{\xi}}$ en $(E^F, \bigotimes_{s \in F} \mathcal{E}_s)$, $(\mathcal{E}_s = \mathcal{E}$, $s \in F)$, inducida por la variable aleatoria $(\xi_s)_{s \in F}$ de $(\Omega, \mathcal{F}, P)$ en $(E^F, \bigotimes_{s \in F} \mathcal{E}_s)$. Se tiene la siguiente *propiedad de compatibilidad*: Si F y F' son subconjuntos finitos y no vacíos de S con $F' \subset F$, entonces $P_{F'}^{\tilde{\xi}} = P_F^{\tilde{\xi}}(p_{F'F})^{-1}$, donde $p_{F'F} : E^F \to E^{F'}$ es la proyección natural.
Al conjunto de probabilidades $\left\{P_F^{\tilde{\xi}} :\ F \text{ subconjunto finito no vacío de } S\right\}$, se le llama *ley de probabilidad* del proceso estocástico $\tilde{\xi}$.

Teorema 6.3. *Sean (X,T) un espacio topológico* **s***-polaco, S un conjunto infinito y para cada F, parte finita no vacía de S, P_F una probabilidad en $(X^F, \bigotimes_{s\in F}(\mathcal{B}(X))_s)$, $((\mathcal{B}(X))_s = \mathcal{B}(X),\ s \in F)$, tal que $P_{F'} = P_F \circ (p_{F'F})^{-1}$ para todo F' y F subconjuntos finitos no vacíos de S con $F' \subset F$. Entonces, existe un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y existe un proceso estocástico $\tilde{\xi} = \{\xi_s\}_{s\in S}$ en $(\Omega, \mathcal{F}, P)$ con dominio de parámetro S y espacio de estados $(X, \mathcal{B}(X))$, con $P_F^{\tilde{\xi}} = P_F$, $F \subset S$, finito no vacío.*

Demostración. Por el **Teorema 4.8**, existe una única probabilidad P en el espacio medible $((\Omega =)X^S, (\mathcal{F} =) \bigotimes_{s\in S}(\mathcal{B}(X))_s)$, donde $(\mathcal{B}(X))_s = B(X)$, $s \in S$, tal que $P \circ (p_{SF})^{-1} = P_F$ para todo subconjunto finito no vacío F de S. Para cada $s \in S$ se tiene la variable aleatoria $(\xi_s =)p_s$ del espacio de probabilidad $(\Omega, \mathcal{F}, P)$ en $(X, \mathcal{B}(X))$, y el proceso estocástico $\tilde{\xi} = \{\xi_s = p_s\}_{s\in S}$ en $(\Omega, \mathcal{F}, P)$ con dominio de parámetro S y espacio de estados $(X, \mathcal{B}(X))$. Finalmente, $P \circ (p_{SF})^{-1} = P_F^{\tilde{\xi}}$. $\square$

Movimiento Browniano

Uno de los ejemplos más importante de los procesos estocásticos, fundamentalmente en las aplicaciones, es el *movimiento Browniano*:
Sean $\tilde{\beta} = \{\beta_t\}_{t\in J_T}$ un proceso estocástico en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ con dominio del parámetro J_T, ($J_T = [0,T]$, si T es un número real positivo, y $J_\infty = [0, +\infty)$), y espacio de estados $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, y r un número real. Se dice que $\tilde{\beta}$ es un *movimiento Browniano (lineal) con origen en* r si,

(1). $P(\beta_0 \neq r) = 0$.

(2). $\tilde{\beta}$ es un proceso estocástico con incrementos independientes y estacionarios, lo cual significa que para todo $t_1 < t_2 < ... < t_n$ en J_T, las variables aleatorias $\beta_{t_2} - \beta_{t_1}, ..., \beta_{t_n} - \beta_{t_{n-1}}$ son independientes, (para todo $B_1, ..., B_{n-1} \in \mathcal{B}(\mathbb{R})$, los sucesos $(\beta_{t_2} - \beta_{t_1})^{-1}(B_1)$,...,$(\beta_{t_n} - \beta_{t_{n-1}})^{-1}(B_{n-1})$, del espacio de probabilidad $(\Omega, \mathcal{F}, P)$, son independientes), y $P(\beta_t - \beta_s \in A) = P(\beta_{t+\delta} - \beta_{s+\delta} \in A)$ para todo s, t, δ con $s, t, s+\delta, t+\delta \in J_T$, $s < t$, $\delta > 0$, y todo $A \in \mathcal{B}(\mathbb{R})$.

(3). Los incrementos $\beta_t - \beta_s$ son variables aleatorias Gaussianas con $E(\beta_t - \beta_s) = 0$ y $V(\beta_t - \beta_s) = \sigma^2|t-s|$, (es decir,
$P(\beta_t - \beta_s \leqslant x) = \int_{-\infty}^{x}(1/\sqrt{2\pi\sigma^2|t-s|})\exp(-u^2/(2\sigma^2|t-s|)du,\ x \in \mathbb{R})$.

(4). Las trayectorias de $\tilde{\beta}$ son continuas.

Si $r = 0$ y $\sigma^2 = 1$ se dice que $\tilde{\beta}$ es un movimiento Browniano estándar.
Sea $\tilde{\beta} = \{\beta_t\}_{t\in J_T}$ un movimiento Browniano estándar en el espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Al proceso estocástico $\tilde{\xi} = \{\xi_t\}_{t\in J_T}$,

$$\xi_t = x_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) \cdot t + \sigma\beta_t\right),\ t \in J_T,$$

donde μ, σ y x_0 son números reales con $\sigma \neq 0$ y $x_0 > 0$, se le llama *movimiento Browniano geométrico.*

La existencia del movimiento Browniano estándar fue probado por R. Wiener en 1923, y posteriormente se demostró que se puede obtener como límite de paseos aleatorios, lo cual justifica, de alguna forma, que dicho movimiento se tome para modelizar el desplazamiento de partículas suspendidas en un fluido (un líquido o un gas).
Sea $\tilde{\xi} = \{\xi_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Supongamos que $E(\xi_n) = 0$ y $V(\xi_n) = 1$ para todo $n \in \mathbb{N}$. Consideramos el paseo aleatorio generado por la sucesión de variables aleatorias dada, es decir, la sucesión de variables aleatorias $\eta_n = \sum_{k=0}^{n} \xi_k$, $n \in \mathbb{N}$, (a las variables aleatorias ξ_n se les llama *pasos* del paseo aleatorio y a las sumas η_n la *posición* del paseo aleatorio después de n *pasos*). Interpolamos linealmente entre los puntos enteros de $\mathbb{R}$, es decir, $\eta_t = \eta_{[t]} + (t - [t])(\eta_{[t]+1} - \eta_{[t]})$, $t \in J_\infty$, (donde $[t]$ es la parte entera de t). De esta forma se obtiene un proceso estocástico con trayectorias continuas, $\{\eta_t\}_{t\in J_\infty}$, en $(\Omega, \mathcal{F}, P)$ con dominio del parámetro J_∞ y espacio de estados $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
Ahora consideramos la sucesión de procesos estocásticos, $\{\tilde{\eta}_n\}_{n\in\mathbb{N}^+}$, donde $\tilde{\eta}_n = \{\eta_{nt}/\sqrt{n}\}_{t\in J_1}$. Por la **Proposición 6.2(2)**, $\psi_{\tilde{\eta}_n}$, $n \in \mathbb{N}^+$, es una variable aleatoria de $(\Omega, \mathcal{F})$ en $(C([0,1], \mathbb{R}), \mathcal{B}(C([0,1], \mathbb{R})))$. Con estas notaciones:

Teorema 6.4 (**Principio de invariancia de Donsker**). *La sucesión de variables aleatorias, $\{\psi_{\tilde{\eta}_n}\}_{n\in\mathbb{N}^+}$, del espacio de probabilidad $(\Omega, \mathcal{F}, P)$ en el espacio medible $(C([0,1], \mathbb{R}), \mathcal{B}(C([0,1], \mathbb{R})))$, converge en distribución a la variable aleatoria $\psi_{\tilde{\beta}}$ de $(\Omega, \mathcal{F}, P)$ en $(C([0,1], \mathbb{R}), \mathcal{B}(C([0,1], \mathbb{R})))$, donde $\tilde{\beta} = \{\beta_t\}_{t\in[0,1]}$ es un movimiento Browniano estándar.*

Los detalles de la demostración de este teorema pueden verse en el capítulo 5 de [7], y en esa demostración se utiliza el **Teorema 5.3**, (la caracterización de la convergencia en distribución por cerrados del espacio topológico considerado), que se aplica en el espacio topológico polaco $(C([0,1], \mathbb{R}), \mathcal{B}(C([0,1], \mathbb{R})))$.
El teorema anterior es uno de los múltiples resultados que ilustran el interés del estudio de las variables aleatorias en espacios de probabilidades definidos en espacios topológicos más generales que los Euclideanos.

Referencias

[1] P. Billingsley: *Convergence of Probability Measures*. Wiley, New York 2009.

[2] J. Margalef Roig, S. Miret Artés, E. Outerelo Domínguez: *Probabilidad y Economía 3. Procesos Estocásticos*. Sanz y Torres, Madrid 2014.

[3] J. Margalef Roig, E. Outerelo Domínguez: *Probabilidad y Economía 1. Mercados Financieros Finitos*. Sanz y Torres, Madrid 2010.

[4] J. Margalef Roig, E. Outerelo Domínguez: *Probabilidad y Economía 2. Espacios de Probabilidad Generales.* Sanz y Torres, Madrid 2013.

[5] J. Margalef Roig, E. Outerelo Domínguez, J. L. Pinilla Ferrando: *Topología, V. 2.* Alhambra, Madrid 1979.

[6] J. Margalef Roig, E. Padrón Fernández, E. Outerelo Domínguez: *Topología General.* Sanz y Torres, Madrid 2000.

[7] P. Mörters, Y. Peres: *Brownian Motion.* Cambridge Series in Statistical and Probabilistic Mathematics, 2010.

Compact Hausdorff group topologies for the additive group of real numbers

Elena Martín-Peinador* **and Montserrat Bruguera Padró**

Departmento de Geometría y Topología
Universidad Complutense de Madrid
28040 Madrid, Spain
peinador@mat.ucm.es

Departament de Matemàtiques
Universitat Politècnica de Catalunya
08028 Barcelona, Spain
m.montserrat.bruguera@upc.edu

To José María Montesinos with thanks for his friendship, and for his willingness to share with us his own approach to beautiful Mathematics.

ABSTRACT

We deal with an example of a topology ν for the additive group of real numbers $\mathbb{R}$, which makes it a compact Hausdorff topological group. Further, $(\mathbb{R}, \nu)$ is connected, but neither arcwise connected, nor locally connected. Thus, it is neither a Lie group, nor a curve in the sense of H. Mazurkiewicz. The contribution of this short note is to provide an elementary proof of the fact that it is not arcwise connected.

2010 Mathematics Subject Classification: 54C40, 14E20
Key words: The real line, topological group, arcwise connected, curve.

*Author partially supported by the project of the Spanish Ministerio de Economía y Competitividad. Grant: MTM 2013-42486.

1. Introduction

The set $\mathbb{R}$ of real numbers is a corner stone in Mathematics. It supports all kind of structures: the simple operations of sum and multiplication make $\mathbb{R}$ the perfect paradigm of a group, a ring or a vector space. If it is endowed with the absolute value it becomes the model for metric spaces and consequently for topological spaces. We could point out so many features of the set $\mathbb{R}$ that make it a model, that maybe the following question is pertinent: is $\mathbb{R}$ a human creature, or is it just a gift of God to Mankind?

In this short note we do not pretend to answer that philosophical question, we only try to report on a concrete structure on $\mathbb{R}$ which is suprising according to our standard intuition. Namely, the additive group of real numbers $\mathbb{R}$ can be endowed with a topology ν such that $(\mathbb{R}, \nu)$ is a compact, connected, Hausdorff topological group which is neither arcwise connected, nor locally connected. Thus $(\mathbb{R}, \nu)$ is neither a Lie group, nor a curve in the sense of Hahn-Mazurkiewicz.

The mentioned topology ν called our attention from a one-page paper of Halmos, and we have denominated it Halmos topology. The route to define it in [7] is first to establish an algebraic isomorphism between $\mathbb{R}$ and the group of all characters on the rational numbers, say $\mathrm{Hom}(\mathbb{Q}, \mathbb{T})$.Taking into account that $\mathrm{Hom}(\mathbb{Q}, \mathbb{T})$ has a natural topology as a subspace of the product $\mathbb{T}^{\mathbb{Q}}$, a topology on $\mathbb{R}$ can be defined under the constrain of making topological (i.e. continuous and open) the algebraic isomorphism. This is precisely what we call the Halmos topology ν on $\mathbb{R}$. By its very definition, the properties of $(\mathbb{R}, \nu)$ are precisely the same as those of the compact group $\mathrm{Hom}(\mathbb{Q}, \mathbb{T})$, and for this reason we devote the first section of this paper to study the latter.

In [11] we dealt with the group $(\mathbb{R}, \nu)$, giving there the details of the definition and a proof that it is not arcwise connected which seemed simple, but a deep artillery was hidden in it. In the present paper we provide an elementary proof of the fact that it is not arcwise connected. We will also give the reference to prove that it is not locally connected.

Notation. All groups considered in this note will be abelian. For a set A we denote by $\mathrm{Card}(A)$ or by $|A|$ the cardinality of A. The symbols $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{T}$ and $\mathfrak{c}$ denote, respectively, the integers, the rational numbers, complex numbers of modulus one, and the cardinality of $\mathbb{R}$. We write $\mathbb{Q}_d$ or $\mathbb{Q}_u$ if $\mathbb{Q}$ must be considered with the discrete topology or with the induced from the euclidean topology of $\mathbb{R}$, and the same meaning have the subindex d and u in other contexts.

Let G, Y be groups. We denote by $\mathrm{Hom}(G, Y)$ the set of all group homomorphisms from G to Y: with operation defined pointwise it becomes a group. If G, Y are topological groups, $\mathrm{CHom}(G, Y)$ stands for the continuous elements of $\mathrm{Hom}(G, Y)$. If $Y = \mathbb{T}$, the elements of $\mathrm{Hom}(G, \mathbb{T})$ are called characters of G, and the group of continuous characters $G^\wedge := \mathrm{CHom}(G, \mathbb{T})$ is called the dual group of G, or the character group of G. Whenever $\mathbb{R}$ or $\mathbb{T}$ are target groups, say in $\mathrm{CHom}(\mathrm{G}, \mathbb{R})$ or $\mathrm{CHom}(\mathrm{G}, \mathbb{T})$, they are supposed to carry the corresponding euclidean topology. Observe that $\mathrm{Hom}(\mathrm{G}, \mathbb{T})$ is a closed subgroup of $\mathbb{T}^G$, therefore it is compact and Hausdorff as a subspace of it. However, $\mathrm{CHom}(\mathrm{G}, \mathbb{T})$ may not be closed in $\mathbb{T}^G$.

All the dual groups considered are supposed to carry **the compact open topology**. This is in consonance with Pontryagin duality theory. Obviously if the original group G is discrete, then $G^\wedge = \mathrm{Hom}(G, \mathbb{T})$ and the compact open topology coincides with the pointwise convergence topology which is precisely the induced by the natural embedding of $G^\wedge$ in the product $\mathbb{T}^G$.

2. Groups of homomorphisms defined on $\mathbb{Q}$

We shall look at $\mathrm{Hom}(\mathbb{Q}, \mathbb{T})$ as the character group of $\mathbb{Q}_d$, the discrete group of rationals. Thus we write $\mathbb{Q}_d^\wedge := \mathrm{Hom}(\mathbb{Q}, \mathbb{T})$, and as said above, it carries the pointwise convergence topology. We also consider $\mathbb{Q}_u^\wedge := \mathrm{CHom}(\mathbb{Q}_\mathrm{u}, \mathbb{T})$ the character group of $\mathbb{Q}$ equipped with the topology induced from the euclidean topology of $\mathbb{R}$. If the target group $\mathbb{T}$ is substituted by $\mathbb{R}$, we have $\mathrm{Hom}(\mathbb{Q}, \mathbb{R})$, which happens to be the same as $\mathrm{CHom}(\mathbb{Q}_u, \mathbb{R})$, as proved below.

2.1. A few elementary facts about $\mathrm{Hom}(\mathbb{Q}, \mathbb{R})$

The set $\mathrm{Hom}(\mathbb{Q}, \mathbb{R})$ is quite simple, since every element $\chi \in \mathrm{Hom}(\mathbb{Q}, \mathbb{R})$ can be identified to the real number $\chi(1)$, as we claim in the following Lemma.

Lemma 2.1 *Every group homomorphism $\chi : \mathbb{Q} \to \mathbb{R}$ is defined by its value in $1 \in \mathbb{Q}$, being χ a linear mapping if $\mathbb{Q}$ and $\mathbb{R}$ are considered as vector spaces over $\mathbb{Q}$. Thus,* $\mathrm{CHom}(\mathbb{Q}_\mathrm{u}, \mathbb{R}) = \mathrm{Hom}(\mathbb{Q}, \mathbb{R})$.

The validity of Lemma 2.1 derives from the facts that $\mathbb{R}$ is divisible and χ an homomorphism: if $\chi(1) = r \in \mathbb{R}$, and m is any integer number, $\chi(1/m)$ must be r/m so that $\chi(1/m + \overset{m}{\ldots} + 1/m) = r$. Thus $\chi(n/m) = (n/m)\chi(1)$, and χ is in fact a linear mapping from the one-dimensional vector space $\mathbb{Q}$ to $\mathbb{R}$ (as a vector spaces over $\mathbb{Q}$). Consequently, $\chi : \mathbb{Q} \to \mathbb{R}$ is continuous if both $\mathbb{Q}$ and $\mathbb{R}$ are endowed with the euclidean topology.

However, if $\phi : \mathbb{Q} \to \mathbb{T}$ is a homomorphism and $\phi(1) \in \mathbb{T}$ is known, two possibilities arise for $\phi(1/2)$, the two square roots of $\phi(1)$ and roughly speaking, many more for $\phi(1/m)$ if $m > 2$. This explains the following negative claim:

Remark 2.1 *A homomorphism $\phi : \mathbb{Q} \to \mathbb{T}$ is not defined by its value in 1. Later on we shall prove that* $\mathrm{Hom}(\mathbb{Q}, \mathbb{T}) \neq \mathrm{CHom}(\mathbb{Q}_\mathrm{u}, \mathbb{T})$.

2.2. The group $\mathrm{Hom}(\mathbb{Q}, \mathbb{T})$.

In order to understand how the set $\mathrm{Hom}(\mathbb{Q}, \mathbb{T})$ looks like, we include the following description done in [8]. Let $\chi : \mathbb{Q} \to \mathbb{T}$ be a homomorphisms from $\mathbb{Q}$ to $\mathbb{T}$. It gives rise to a sequence $\{\alpha_n\}$ of elements of $\mathbb{T}$, namely $\alpha_n := \chi(1/n!)$. Clearly it holds:

$$(\alpha_n)^n = \alpha_{n-1}. \ (*)$$

On the other hand, every sequence $\{\beta_n\} \subset \mathbb{T}$ which satisfies the condition (*) gives rise to a character ϕ on $\mathbb{Q}$. Just define $\phi(1/n!) = \beta_n$, and take into account the expression of $\mathbb{Q}$ as a union of cyclic groups, $\mathbb{Q} = \bigcup_n (1/n!)$.

Thus, the character group of $\mathbb{Q}_d$ can be identified to a limit of an inverse sequence $\{\mathbb{T}, g_n\}$, where the linking mappings $g_n : \mathbb{T} \to \mathbb{T}$ are defined by $g_n(\gamma) = \gamma^n$.

2.3. $\mathrm{Hom}(\mathbb{Q}, \mathbb{T}) \neq \mathrm{CHom}(\mathbb{Q}_\mathrm{u}, \mathbb{T})$.

An example of a character χ defined on $\mathbb{Q}$ which is not continuous considered as a mapping from $\mathbb{Q}_u$ into $\mathbb{T}$ is provided in [5]. Following 2.2, it is achieved by means of a sequence $\{\alpha_n\}$ for which (*) holds but α_n does not converge to $1 \in \mathbb{T}$. The character χ defined through $\chi(1/n!) = \alpha_n$ for such an $\{\alpha_n\}$ cannot be continuous if $\mathbb{Q}$ carries the euclidean topology. In fact, from $1/n! \to 0$ in $\mathbb{Q}_u$ we would obtain $\alpha_n \to \chi(0) = 1$, which does not happen. We present the concrete example below.

Example 2.1 *Let $\alpha_n := e^{2\pi i \theta_n}$, being*

$$\theta_n = \frac{1}{2(n!)} + \frac{1}{n!}\left(\sum_{i=2}^{n} \left[\frac{i}{2}\right](i-1)!\right),$$

where the symbol $[.]$ stands for entire part. From the equality:

$$\theta_n = \frac{\theta_{n-1} + [n/2]}{n}, \ \textit{for all } n \in \mathbb{N} \setminus \{0, 1\},$$

it is straightforward to check that $\alpha_n^n = \alpha_{n-1}$, therefore χ is a character on $\mathbb{Q}$. On the other hand, it can be proved by induction that

$$1/4 \leq \theta_n \leq 3/4, \ \textit{for all } n \in \mathbb{N} \setminus \{0, 1\},$$

which implies $\alpha_n \not\to 1$ in $\mathbb{T}$.

3. Lifting continuous characters of a topological group G to real continuous characters

Taking into account that $\mathbb{R}_u$ is the universal cover of $\mathbb{T}$ by means of the exponential mapping ρ, it is natural to think about the possibility of lifting continuous characters defined on a topological group G, to continuous homomorphisms from G to $\mathbb{R}$. First a definition to make precise what we mean.

Definition 3.1 *Let G be a topological group. A continuous character $\phi : G \to \mathbb{T}$ is said to be liftable over the reals if there exists a continuous homomorphism $\tilde{\phi} : G \to \mathbb{R}$ such that $\rho\tilde{\phi} = \phi$. The term real character is used for a homomorphism from G to $\mathbb{R}$, and we will call $\tilde{\phi}$ a lift of ϕ.*

Let G be a topological group. Does the statement "every continuous character on G is liftable" hold?

Clearly for a simply connected group G every continuous character is liftable. (This follows from well known properties of the theory of coverings.) However, one of the most elementary connected groups, like $\mathbb{T}$ is not simply connected. Thus, the assumption of simple connectedness is too strong. Nevertheless, the statement will hold for any topological real vector space considered as a group.

Partial Answers 3.1 *1)* (Dixmier [6]) *If G is a locally compact group, and $G^\wedge$ is arcwise connected then every continuous character is liftable.*

2) If G is a k-space and $G^\wedge$ is arcwise connected, then every continuous character is liftable.

Obviously the result 1) is a particular case of 2) and 2) can be proved by means of the homotopy lifting property of a covering [14, Chapter 2, thm. 3]. Dixmier gives a proof of 1) ad hoc for locally compact groups, based upon the Structure Theorem for the latter, which is a delicate matter. The assertion 2) is from [12]. A detailed proof of it can be seen in [3, 5.2.1 and 5.2.2].

The possibility of lifting the characters of a topological group G to real characters by requiring conditions on the arcs of $G^\wedge$, rather than the global property of arcwise connectivity, is analyzed in a series of papers that have recently appeared. As defined in [2], a topological group G has the EAP (equicontinuous arc property) if every arc in $G^\wedge$ is equicontinuous with respect to the original topology of G. In the mentioned paper, the set of continuous characters of G that can be lifted is called $G^\wedge_{\text{lift}}$. It is a subgroup of $G^\wedge$ contained in the arc component, denominated $G^\wedge_a$. The main result of [2] asserts that $G^\wedge_{\text{lift}} = G^\wedge_a$ whenever G has the EAP property.

Since we have extensively dealt with the continuous convergence structure Λ defined on the dual of a topological group (see for instance [4] and [3]), it is clear for us that the EAP property has a natural setting in terms of it. Let us state here how to reformulate the result of [2] quoted above:

Theorem 3.2 *Let G be a topological abelian group. The following assertions are equivalent:*

1) Every character $\phi \in G^\wedge$ can be lifted over the reals.

2) The group $\mathrm{CHom}(G, \mathbb{T})$ *endowed with the continuous convergence structure Λ is arcwise connected.*

4. The main results

In order to obtain a compact Hausdorff group topology on $\mathbb{R}$, we first state the following:

Proposition 4.1 *The group $\mathbb{Q}_d^\wedge$ is algebraically isomorphic to the group of real numbers $\mathbb{R}$.*

A detailed proof of this assertion can be seen in [11]. For the readers convenience we point out that it derives from the fact that $\mathbb{Q}_d^\wedge$ and $\mathbb{R}$ are vector spaces over the field of rationals $\mathbb{Q}$, such that $|\mathbb{Q}_d^\wedge| = |\mathbb{R}|$. Their respective Hamel bases have also the same cardinality. Thus, any fixed bijection from a Hamel basis of $\mathbb{R}$ to a Hamel basis of $\mathbb{Q}_d^\wedge$ can be extended to an algebraic isomorphism. For definiteness, let us call $\phi : \mathbb{R} \to \mathbb{Q}_d^\wedge$ such an isomorphism, and let us denote by ν the topology on $\mathbb{R}$ that makes ϕ a **topological** isomorphism. We call ν a **Halmos topology** for $\mathbb{R}$.

Since $\mathbb{Q}_d^\wedge$ and $(\mathbb{R}, \nu)$ are isomorphic topological groups, we prove now that the first one is not arcwise connected in order to have the same property for $(\mathbb{R}, \nu)$.

Theorem 4.2 *The group $\mathbb{Q}_d^\wedge$ is not arcwise connected.*

Proof. Assume by contradiction that $\mathbb{Q}_d^\wedge$ were arcwise connected. Since $\mathbb{Q}_d$ is locally compact by the assertion 1) in 3.1, every character on $\mathbb{Q}$ would be liftable over the reals. In particular χ, as described in the Example would be liftable to a real character $\tilde{\chi} : \mathbb{Q}_d \to \mathbb{R}$ such that $\rho\tilde{\chi} = \chi$. By the equality $\mathrm{CHom}(\mathbb{Q}_\mathrm{u}, \mathbb{R}) = \mathrm{Hom}(\mathbb{Q}, \mathbb{R})$ (see 2.2), we would have that $\tilde{\chi}$ is continuous with respect to the usual topology. But in that case $\rho\tilde{\chi} = \chi$ would be continuous with respect to the usual topology of $\mathbb{Q}$, in other words $\chi = \rho\tilde{\chi} \in \mathrm{CHom}(\mathbb{Q}_\mathrm{u}, \mathbb{T})$, which does not hold as proved in the Example. $\square$

We provide now references which easily prove the facts that $(\mathbb{R}, \nu)$ is neither locally connected, nor a Lie group.

Proposition 4.3 [13, Theorem 42, pp. 169] *A compact locally connected and connected abelian group decomposes into the direct sum of a finite or countable number of subgroups, each isomorphic with the group $\mathbb{T}$.*

Clearly our group $(\mathbb{R}, \nu)$ cannot have any subgroup isomorphic to $\mathbb{T}$, since in $\mathbb{R}$ there are no torsion elements. Because of the definition of the topology ν, all the topological assertions done for $\mathbb{Q}_d^\wedge$ hold for $(\mathbb{R}, \nu)$, thus we already know that $(\mathbb{R}, \nu)$ is compact and connected. Therefore, according to Proposition 4.3, it cannot be locally connected. On the other hand a compact Lie group is necessarily locally connected (see, for instance [13, Remark H, p. 212]).

Finally we recall the following:

Proposition 4.4 [9, Theorem H. Mazurkiewicz] *A Hausdorff topological space is a curve if and only if it is a metrizable Peano continuum.*

By a Peano continuum it is meant a Hausdorff, compact, connected and locally connected topological space. We now have all the ingredients to formulate the Theorem:

Theorem 4.5 *The topological group* $(\mathbb{R}, \nu)$ *is compact, Hausdorff and connected, but neither arcwise connected nor locally connected. Therefore* $(\mathbb{R}, \nu)$ *is neither a Lie group, nor a curve in the sense of H. Mazurkiewicz.*

References

[1] L. Aussenhofer: *On the arc component of a locally compact abelian group.* Math. Z. **257** (2) (2007) 239–250.

[2] L. Aussenhofer, M. J. Chasco, X. Domínguez: *Arcs in the Pontryagin dual of a topological abelian group.* J. Math. Anal. Appl. **425** (1) (2015) 337–348

[3] M. Bruguera: *Grupos topológicos y grupos de convergencia: estudio de la dualidad de Pontryagin.* Doctoral Dissertation. Barcelona, 1999.

[4] M. J. Chasco, E. Martín Peinador: *Binz-Butzmann duality versus Pontryagin duality.* Arch. Math. (Basel) **63** (3) (1994) 264–270.

[5] E. Cillero: *Caracteres del grupo aditivo de los racionales.* Rev. Real Academia de Ciencias de Zaragoza. **58** (2003) 115–128.

[6] M. J. Dixmier: *Quelques propriétés des groupes abéliens localement compacts.* Bull. Sci. Math. **81** (2) (1957) 38–48.

[7] P. R. Halmos: *Comment on the real line.* Bull. Amer. Math. Soc. **50** (1944) 877–878.

[8] E. Hewitt, K. A. Ross: *Abstract Harmonic Analysis.* Vol. I . Grundlehren der mathematischen Wissenschaften **115.** Springer-Verlag, Berlin 1963.

[9] J. Hocking, G. Young: *Topology.* Addison-Wesley, Massachusets 1961.

[10] K. H. Hofmann, S. A. Morris: *The Structure of Compact Groups.* de Gruyter Studies in Mathematics **25.** Walter de Gruyter, Berlin - New York 1998.

[11] E. Martín-Peinador, C. T. Stevens: *Completeness properties of group topologies for* $\mathbb{R}$. Topology Appl. **192** (2015) 169–175.

[12] P. Nickolas: *Reflexivity of topological groups.* Proc. Amer. Math. Soc. **65** (1) (1977) 137–141.

[13] L. Pontrjagin: *Topological groups.* Princeton University Press, Princeton Mathematical Series **2** 1946.

[14] E. H. Spanier: *Algebraic Topology.* McGraw-Hill, New York 1966.

Unraveling the Dogbone Space

Diego Mondéjar Ruiz

Departamento de Geometría y Topología
Universidad Complutense de Madrid
28040 Madrid, Spain
dmondejar@mat.ucm.es

Dedicado con cariño y admiración al profesor José María Montesinos Amilibia en su jubilación. Gracias José, por enseñarme tanto.

ABSTRACT

The aim of this paper is to recall an upper semicontinuous decomposition of $\mathbb{R}^3$ discovered by R. H. Bing, called the Dogbone space, that enjoy very interesting and non-intuitive properties and to scketch, emphasizing the visualization of the procedures, a proof due to Bing of a remarkable theorem concerning it. We also prove a theorem of Shapiro about the Whitehead manifold using similar arguments.

2010 Mathematics Subject Classification: 54B15, 57M40, 57N10, 57N37.
Key words: Upper semicontinuous decomposition, shrinkability, manifold factors, Dogbone space, Whithead manifold.

1. Introduction

When we deal with four dimensional spaces, our intuitions, coming from the three dimensions into which we exist, turn frustrated. The extra dimension allows us to do constructions that escape to our understanding, producing amazing results. It is a well known result that every knot can be untied in $\mathbb{R}^4$. Altough we cannot imagine it, there is a homeomorphism of $\mathbb{R}^4$ moving continuously the knot to a standard circle. The idea is that, for every crossing of the knot, we can change a little bit the fourth coordinate of one of the components and move it without intersecting the other, to exchange them. There is enough space to achieve this. So knot theories in $\mathbb{R}^4$ does not make any sense. Here, we are concerned with some pathological spaces, defined in the three-dimensional space, that produce surprising results when considered in $\mathbb{R}^4$.

In 1935, Whitehead gave a failed proof of the Poincaré Conjecture for open 3-manifolds. He "showed" [16] that the only open contractible 3-manifold is the three dimensional sphere $\mathbb{S}^3$. After he noticed his own mistake, he constructed [17] the first example of an open 3-manifold that is contractible but is not homeomorphic to $\mathbb{R}^3$, later called the Whitehead manifold. Some years after, Shapiro observed that the cartesian product of the Whitehead manifold with the real line was homeomorphic to $\mathbb{R}^4$. At that time, R. H. Bing had discovered a striking example in decomposition theory, the Dogbone decomposition space [5]. This space is defined as a quotient of a fairly complicated decomposition, and Bing had shown that it was different to $\mathbb{R}^3$. But he was decided to understand the complexity of the linkage of the elements of such a decomposition. It was, thinking about the result of Shapiro, when Bing discovered that he could manipulate, in a very smart way, the elements of the decomposition space in $\mathbb{R}^4$ to obtain that the cartesian product of the Dogbone space with the real line is homeomorphic to $\mathbb{R}^4$ [6], the result that we are going to analyze here. These results were the first showing that $\mathbb{R}^4$ has non trivial factors. Afterwards, there was a series of papers with generalizations to more dimensions and kinds of spaces, for example, [2] and [15]. It is in Bing own words [7], where we find that these results are shocking: "We are learning strange things about cartesian products".

I wish to thank professor Montesinos for introducing me to this beautiful topic and for his inspirational comments and advices, either in a classroom or in a couch.

2. Decomposition spaces and shrinkability

Decomposition space theory formalizes the idea of crushing some parts of a topological space. Given a topological space, we can select a partition into disjoint sets and consider the quotient of the space by these sets. A *decomposition* of a topological space X is a collection D of mutually exclusive sets covering X. We call the elements of D that are just points of X the *degenerate* elements of the decomposition, and the rest are called *non-degenerate*. Given any decomposition of X, we can consider the topological quotient $\mathcal{D} = X/D$, called the *decomposition space*, whose points are the elements of the decomposition D, and has as open sets collections of elements of D contained in an open set of X. A decomposition of X is said to be *upper semicontinuous* if, for every $d \in D$ and every open set $U \subseteq X$ containing d, there exists an open set V containing d such that every d' intersecting V is contained in U. This condition comes from topological considerations about the behaviour of the quotient regarding the initial space (for example Hausdorfness, metrizability, etc). As a general reference about decomposition spaces, we recommend [8].

We are mainly interested in decompositions such that the elements of the decomposition have some nice properties. If the elements of an upper semicontinuous decomposition are compact continua, it is called a *monotone* decomposition. In particular, there is a lot of work about decompositions of euclidean spaces. We are interested to know when a monotone decomposition of an euclidean space $\mathbb{R}^n$ yields a decomposition space homeomorphic to $\mathbb{R}^n$. As in many cases in topology, the problem is relatively easy for dimension 2 but it turns out extremely complicated in dimension 3. For the euclidean plane, we have the following complete theorem of R. L. Moore about decomposition spaces:

Theorem 2.1 *(Moore [14]) Every monotone decomposition D of $\mathbb{R}^2$ such that no element of D separates the plane yields a decomposition space $\mathbb{R}/D$ homeomorphic to $\mathbb{R}^2$.*

Roughly speaking, Moore characterized the euclidean plane with some topological properties and showed that any decomposition satisfying the conditions of the theorem gives a decomposition space with the necessary topological properties to be $\mathbb{R}^2$. As we pointed out, the situation becomes much more complicated in dimension 3. It was G. T. Whyburn in [18] who first proposed to find what conditions we need to impose to upper semicontinuous (or monotone) decompositions of $\mathbb{R}^3$ in order to get $\mathbb{R}^3$ as decomposition space. For instance, even for decomposition spaces with only one non-degenerate element, the result is false: The decomposition space with the Fox-Artin arc [11], whose complement in $\mathbb{R}^3$ is not simply connected, as the only non-degenerate element of the decompositon, is not homeomorphic to $\mathbb{R}^3$. Whyburn conjectured that a good candidate of such a condition was that the elements of the decomposition have complement in $\mathbb{R}^3$ homeomorphic to the complement of a point. This property was later called to be *point-like*. With this condition, Moore's theorem can be reformulated as: Every point-like monotone decomposition of $\mathbb{R}^2$ gives a decomponsition space homeomorphic to $\mathbb{R}^2$. The Dogbone space, introduced by Bing (see section 3 below), is a counterexample to this conjecture. Since then, there has been a lot of work around conditions on monotone decompositions of $\mathbb{R}^3$ yielding $\mathbb{R}^3$ as decomposition space. See [8] or the survey [3] for more information.

One method of showing that a decomposition space is homeomorphic to the original space is to continuously deform or shrink the nondegenerate elements of the decomposition to smaller sets, in a controlled way, without disturbing too much the rest of them. This naive (but not easy to carry out) idea was first used by R. H. Bing and is usually called shrinkability. The first use of the shrinkability method is probably in [4], where Bing showed that the union of two solid horned spheres [1] glued along their corresponding points in the boundary is homeomorphic to $\mathbb{S}^3$. This result is very important, but also the method of proving it was a breakthrough in geometric topology: Bing constructed a point-like decomposition space of $\mathbb{R}^3$ that was the union of two solid horned spheres and then he shrunk the elements of the decomposition to show that the decomposition space was actually homeomorphic to $\mathbb{R}^3$. The shrinkability method started to be used in many papers and in several forms and it led to what today is known as Bing's shrinking criterion. There are different versions and generalizations for spaces satisfying some topological properties. The idea is to find conditions that allow us to shrink elements of a decomposition into points. Then, under some hypothesis over the original space we can conclude that the decomposition space is homeomorphic to the original space. In [8], there is a reasonable summary about shrinkability. The shrinkability method has produced a great number of important results in geometric topology, as the proof of the generalized Schoenflies theorem [9] by Morton Brown.

For the sake of simplicity, we will only describe a shrinkability method for decompositions of $\mathbb{R}^n$ useful for our purposes. This is the version used by Bing in [6]. Sometimes, the homeomorphism between the original space and the decomposition space can be accomplished as the limit of some shrinking homeomorphisms. Using pseudo-isotopies is a way of obtaining them. A *pseudo-isotopy* of $\mathbb{R}^n$ is a continuous map $f : \mathbb{R}^n \times [0, 1] \to \mathbb{R}^n$ such that, for every $t \in [0, 1)$, $f_t = f(\cdot, t) : \mathbb{R}^3 \to \mathbb{R}^3$ is a homeomorphism. Given a decomposition D of $\mathbb{R}^n$, we will say that a pseudo-isotopy f *shrinks* the decomposition D, if this three conditions hold:

i) f_0 is the identity map,

ii) for every $t < 1$, f_t is a homeomorphism onto $\mathbb{R}^n$, and

iii) f_1 is onto and sends each element of D onto a different point of $\mathbb{R}^n$.

When we are able to find such a pseudo-isotopy, it is clear that the decomposition space is homeomorphic to $\mathbb{R}^n$. The map $g : \mathbb{R}^n/D \to \mathbb{R}^n$ given by $g([x]) = f_1(x)$ is then a homeomorphism.

3. The Dogbone space

The Dogbone space is a decomposition space of $\mathbb{R}^3$, introduced by R. H. Bing in [5], to show the existence of monotone decompositions of $\mathbb{R}^3$ into points and tame arcs such that there is no pseudo-isotopy shrinking the arcs. Afterwads, he realized that, if the decomposition space was homeomorphic to $\mathbb{R}^3$, then it would be shrinkable by a pseudo-isotopy, so that could not be the case. Hence, that was the first example of a monotone decomposition space of $\mathbb{R}^3$ into point-like compact continua that is not homeomorphic to $\mathbb{R}^3$, a counterexample to the generalization of Moore's theorem 2.1.

The construction is as follows. Consider a solid torus T embedded in $\mathbb{R}^3$ in the standard way. Then, consider, inside of T, four solid tori T_i, $i = 1, \ldots, 4$, embedded exactly as shown in figure 1. Note that the way the solid tori are knotted to each other is the same in both sides: Let us imagine that if we cut the figure transversally to the paper and we rotate the left part π radians, we obtain exactly the right part. We write $X_1 = \{T_1, T_2, T_3, T_4\}$ for the collection of solid tori and $X_1^* = \bigcup_{i=1}^4 T_i$ for the union of its points. Now, in each solid torus T_i, consider another four solid tori T_{ij}, for $j = 1, \ldots, 4$, embedded in the same way that each T_i was embedded in T. We proceed in this way sequentially. In the stage n, we consider four solid tori embedded in each of the 4^{n-1} solid tori considered in the previous stage of the construction,

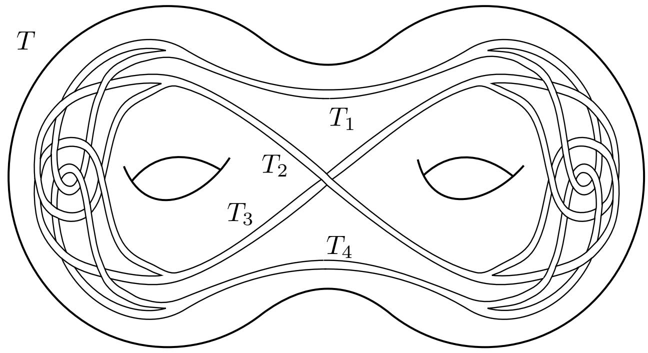


Figure 1: The first stage in the construction of the Dogbone decomposition.

obtaining 4^n solid tori $T_{i_1 i_2 \dots i_n}$. We will denote

$$X_n = \{T_{i_1 \dots i_n} : i_j \in \{1,2,3,4\}\} \text{ and } X_n^* = \bigcup_{T \in X_n} T$$

as the collection of the tori and the union of its points respectively. Continue this process until the infinity, obtaining

$$X_\infty = \{T_{i_1 i_2 \dots} : i_j \in \{1,2,3,4\}\} \text{ and } X_\infty^* = \bigcup_{T \in X_\infty} T.$$

It is clear from the definition that the elements of X_∞ are disjoint tame arcs. In the construction, we can assume that all the solid tori involved are long and sharped enough, such that no element of X_∞ meets in more than one point to any imaginary plane perpendicular to the longest diameter of T. Consider the collection D of subsets d of $\mathbb{R}^3$ such that, d is an element of X_∞ or a point of $\mathbb{R}^3 \setminus X_\infty^*$. Then, D is an upper semicontinuous decomposition of $\mathbb{R}^3$ into points (degenerate elements) and tame arcs (non-degenerate elements). The corresponding decomposition space $\mathcal{D} = \mathbb{R}^3/D$ is called the *Dogbone decomposition space*.

Note that the set of non-degenerate elements of the decomposition D corresponds to a Cantor set in the decomposition space $\mathcal{D}$. If we imagine (again) a plane P intersecting T perpendicular to its longest diameter, we obtain that $X_\infty \cap P$ is a Cantor set in $\mathbb{R}^2$, as shown in figure 2. Each element of X_∞ is a point in the decomposition space $\mathcal{D}$, therefore the non-degenerate elements are a Cantor set in $\mathcal{D}$.

The Dogbone space is an intricate space. Not only it is different from $\mathbb{R}^3$ but it turns out that it is not even a manifold in each of the points of the space corresponding to non-degenerate elements of the decomposition. Intuitively, if one tries to shrink the nondegenerate elements of the decomposition, it seems that we need to stretch others.

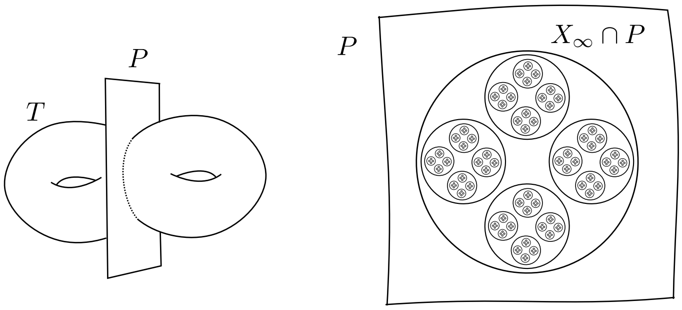


Figure 2: A planar Cantor set.

4. Give me one more dimension and I will unravel the Dogbone space

The Dogbone space is not only known for being the quoted counterexample. It is also the first example of an important fact about cartesian products: The four dimensional euclidean space has non manifold factors. Concretely, Bing showed, in the beautiful paper [6], the following surprising result.

Theorem 4.1 *The cartesian product of the Dogbone space $\mathcal{D}$ with $\mathbb{R}$ is homeomorphic to $\mathbb{R}^4$.*

In order to obtain this result, consider the upper semicontinuous decomposition D' of $\mathbb{R}^4$ consisting of elements $d \times t$ with $d \in D$ and $t \in \mathbb{R}$. The corresponding decomposition space will be $\mathcal{D}' = \mathcal{D} \times \mathbb{R}$. Then, it is possible to find a pseudo-isotopy that shrinks simultaneously the non-degenarate elements of D to points and, hence, the decomposition space $\mathcal{D} \times \mathbb{R}$ is homeomorphic to $\mathbb{R}^4$. So, what it is not possible to do in three dimensions, becomes realizable when we pass to four and take the cartesian product with $\mathbb{R}$. With the additional dimension, we are able to unravel and shrink the elements of the decomposition to points.

The key idea in [6] that makes this shrinkability possible is the following theorem, in which we find useful 4-cells (an n-cell is any topological space homeomorphic to a closed standard unit ball in $\mathbb{R}^n$) containing the non-degenerate elements of the decomposition $\mathcal{D}'$. Then, we can define homeomorphisms fixed outside the 4-cells. We begin with the construction of a 4-cell in the first stage of the decomposition space $\mathcal{D}'$. From now on, when working in $\mathbb{R}^4$, we will call *spatial* coordinates to the first three and *temporal* coordinate to the fourth.

Theorem 4.2 (Bing [5]) *Let T and T_i, $i = 1, \ldots, 4$, be the double solid tori described in the decomposition D of $\mathbb{R}^3$ (depicted in figure 1) and consider any real* *closed interval* $[a,b]$ *and a real number* $\varepsilon > 0$. *There exists a* 4*-cell* K *such that*

$$\bigcup_{i=1}^{4} T_i \times [a,b] \subset K \subset T \times [a-\varepsilon, b+\varepsilon].$$

This theorem means that we can find a 4-cell containing the union of the four double solid tori cross a closed interval but this cell is contained in the product of T with a longer (infinitesimally longer is enough) closed interval.

Proof. (Theorem 4.2) Let us consider four disks D_i, $i = 1,\dots,4$, slicing the solid torus T in three 3-cells C_L, C_M and C_R, as shown in figure 3. It is important that the disks intersect the tori as in the picture, since we need $T_i \cap C_L$, $T_i \cap C_M$ and $T_i \cap C_R$ to be 3-cells. Let us focus on the first two, the central and the left part. For the right part, we will proceed by symmetry. Consider $C_L^+, C_L^- \subset C_L$ the resulting 3-cells by removing an small thickening of the disks D_2 and D_1 respectively (see figure 3). Suppose that the thickenings of the disks are small enough such that C_L^+ and C_L^-

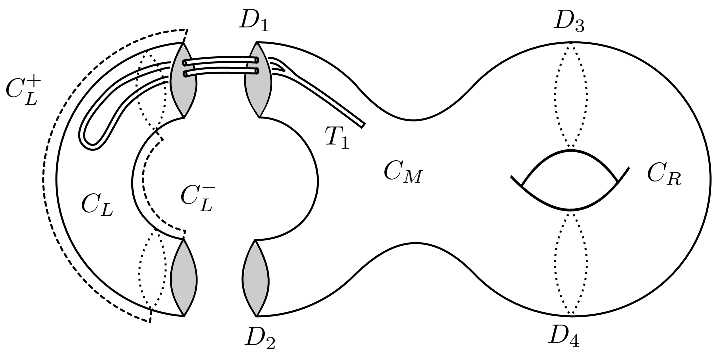


Figure 3: The double torus is cutted into cells.

contains the linking part of the four double solid tori embedded in T, that is,

$$C_L \cap (T_1 \cup T_2) \subset C_L^+ \text{ and } C_L \cap (T_3 \cup T_4) \subset C_L^-.$$

Now, we consider the following seven 4-cells:

$$\begin{aligned}
&C_B = C_M \times [a-\varepsilon, b+\varepsilon],\\
&C_{LW}^+ = C_L^+ \times [a-\varepsilon, a-\frac{\varepsilon}{2}], \quad C_{LW}^- = C_L^- \times [b+\frac{\varepsilon}{2}, b+\varepsilon],\\
&C_{L1} = (C_L \cap T_1) \times [a-\frac{\varepsilon}{2}, b], \quad C_{L2} = (C_L \cap T_2) \times [a-\frac{\varepsilon}{2}, b],\\
&C_{L3} = (C_L \cap T_3) \times [a, b+\frac{\varepsilon}{2}], \quad C_{L4} = (C_L \cap T_4) \times [a, b+\frac{\varepsilon}{2}].
\end{aligned}$$

We claim:

i) The 4-cells C^+_{LW} and C^-_{LW} has empty intersection, since their temporal coordinates never meet.

ii) The intersection of C_B with each of C^+_{LW} and C^-_{LW} are the solid cilinders (hence 3-cells):

$$D_1 \times [a-\varepsilon, a-\frac{\varepsilon}{2}] \text{ and } D_2 \times [b+\frac{\varepsilon}{2}, b+\varepsilon].$$

iii) The union of the three 4-cells $C_B \cup C^+_{LW} \cup C^-_{LW}$ is a 4-cell. Because of *i)* and *ii)*, we are sewing to the 4-cell C_B two mutually exclusive 4-cells along 3-cells. So the union is a 4-cell, that we shall write as C_{BL}. See figure 4 to visualize this cell restricted to some of its dimensions.

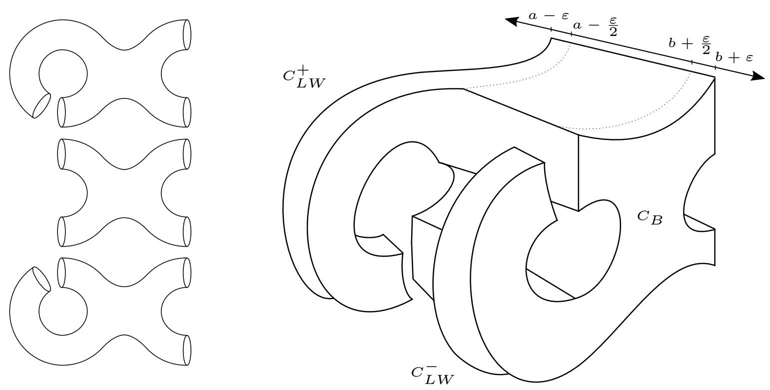


Figure 4: Some views of C_{BL}. In the left part, up to down, the spatial part when the temporal coordinate belongs to $[a-\varepsilon, a-\frac{\varepsilon}{2}]$, $[a-\frac{\varepsilon}{2}, b+\frac{\varepsilon}{2}]$ and $[b+\frac{\varepsilon}{2}, b+\varepsilon]$ respectively. In the right part a representation omitting one spatial coordinate, where the arrow indicates the direction of the temporal coordinate.

iv) The rest of 4-cells, C_{Li} with $i = 1, \ldots, 4$, are mutually disjoint because none of them share any spatial coordinate.

v) The intersection of C_{BL} with each C_{Li} (with $i = 1, \ldots, 4$) is a 3-cell. For example (and the argument holds for the rest of intersections), if we apply the De Morgan's laws to the intersection $C_{BL} \cap C_{L1}$, we obtain

$$C_{BL} \cap C_{L1} = \Big((C_M \cap C_L \cap T_1) \times [a-\frac{\varepsilon}{2}, b]\Big) \cup \Big((C_L \cap T_1) \times \{a-\frac{\varepsilon}{2}\}\Big),$$

which is a 3-cell, as we can see in figure 5. Note that here we see why it is necessary the use of the 4-cells C^+_{LW} and C^-_{LW}, even when they do not contain any part of $\bigcup_{i=1}^{4} T_i \times [a,b]$. The intersection of C_M directly with C_{L1} is not a 3-cell but two disjoint 3-cells (figure 5) and hence the union would not be a 4-cell. So the 4-cells C^+_{LW} and C^-_{LW} acts as wings (hence its notation) or linkage so that the union of C_{BL} with each C_{Li} forms a 4-cell.

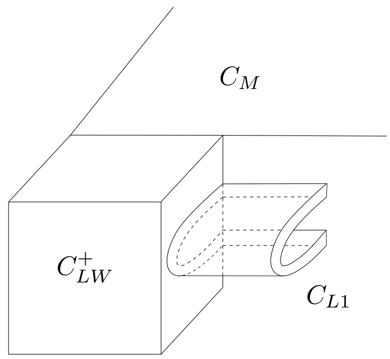


Figure 5: $C_M \cap C_{L1}$ are two 3-cells, but $C_M \cap C^+_{LW} \cap C_{L1}$ is one.

From *i*)-*v*), we can conclude that the union of the seven 4-cells is a 4-cell. Figure 6 shows the structure of unions and intersections between the seven 4-cells. The only caution we must have is the intersection quoted in *v*). Because of the simetry of the Dogbone construction, we can proceed exactly in the same way with the right part, and then we add the six 4-cells $C^+_{RW}, C^-_{RW}, C_{R1}, C_{R2}, C_{R3}, C_{R4}$ to the seven previously written to obtain K, a 4-cell satisfiying the conditions of the theorem. □

At each stage of the decomposition, the solid tori are embedded in the same way T_i are embedded in T, so theorem 4.2 can be actually applied to every solid tori in the construction and we get that, making smaller and smaller the interval of the theorem, we can go as deep as wanted. This is formalized in the following

Corollary 4.1 *Let $T \in X_n$ be a double solid torus of the n-stage in the construction of the dogbone decomposition space. Consider any closed real interval $[a,b]$. For every $t \in \mathbb{N}$ and every partition of $[a,b]$ into $2t$ points $a < a_1 < \ldots < a_s < b_s < \ldots < b_1 < b$, there exist $K_1, K_2, \ldots, K_t$, where K_i is the union of 4^i 4-cells that do not meet each other, such that*

$$\begin{aligned} & T \cap X^*_{n+s} \times [a_s, b_s] \subset Int\, K_s \subset K_s \subset \\ & \subset T \cap X^*_{n+s-1} \times [a_{s-1}, b_{s-1}] \subset Int\, K_{s-1} \subset K_{s-1} \subset \\ & \subset \ldots \subset \\ & \subset T \cap X^*_{n+1} \times [a_1, b_1] \subset Int\, K_1 \subset K_1 \subset \\ & \subset T \times [a,b]. \end{aligned}$$

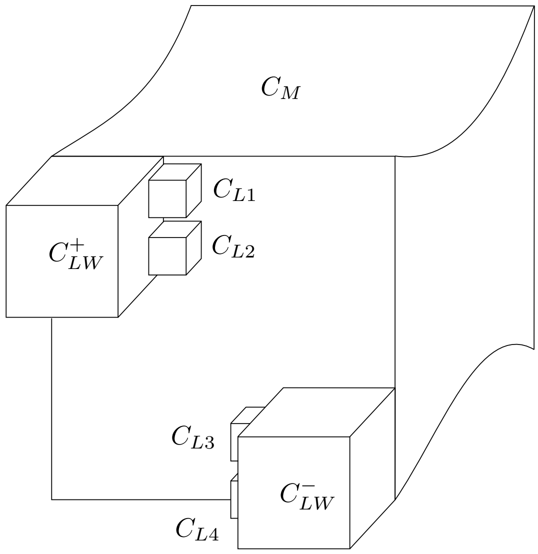


Figure 6: The structure of the 4-cell as union of 4-cells.

For an schematic illustration of this corollary, see figure 7.

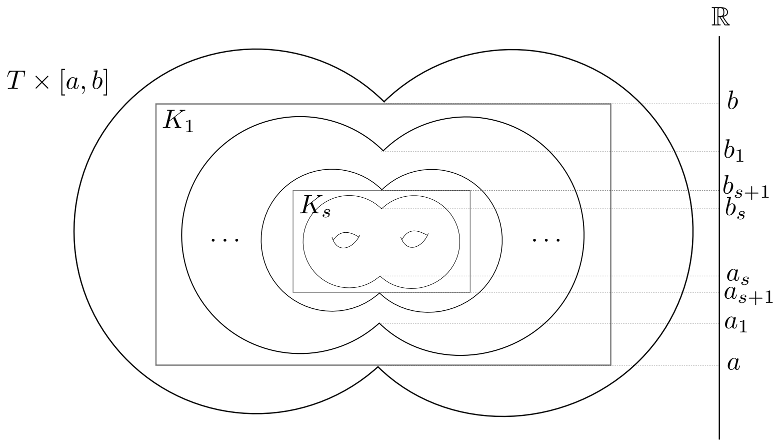


Figure 7: The 4-cells between the tori in several stages.

The rest of the proof of theorem 4.1 is a very technical construction of shrinking homeomorphisms. We give an sketch of the proof. For the details, see the original source [6].

Proof. (sketch, theorem 4.1) Because of theorem 4.2, we can define homeomorphisms of $\mathbb{R}^4$ that are the identity outside a 4-cell and move compact subsets of the interior. We divide the shrinking into four steps:

1. Every double solid torus $T \in X_n$ can be divided into vertical regions. We can define homeomorphisms of $\mathbb{R}^4$ moving to the right part of the interior of the 4-cells that we got in corollary 4.1. We can control to which region we send the spatial part, and, altough we do not know exactly where we are moving the temporal coordinate, we have intervals bounding this move.

2. We fit the bounding intervals in order to do also a symmetric shrinking from the right regions to the left. With this intervals we fill the real line, so we move to the right or to the left depending on the temporal coordinate but everything shrinks to the center. So, we can define a uniformly continuous homeomorphism shrinking $T \in X_n$ to the center.

3. For every $T \in X_n$ and every $\varepsilon > 0$ we can construct a uniformly continuous homeomorphism of $\mathbb{R}^4$ such that it is the identity outside $T \times \mathbb{R}$, it does not move the temporal coordinates more than ε and such that we control the diammeter of the points of deeper tori inside T, after moved by this homeomorphism. We can do this by modifying, according to ε, the scale of the bounding intervals and the number of regions into which we divide T.

4. Finally we take a decreasing sequence of positive real numbers converging to zero and, applying the previous step, we inductively get a sequence of isotopies whose limit is the wanted pseudo-isotopy that shrinks the elements of the decomposition D'.

□

5. Starting point: The Whitehead manifold

The Dogbone space was the first example of a non-manifold factor of $\mathbb{R}^4$. But it was not the first example of non-trivial (i.e., different from $\mathbb{R}^3$) factor of $\mathbb{R}^4$. The first one was the Whitehead manifold. Let us define it. Consider a solid torus T_1 embedded in $\mathbb{S}^3$ (considered as the compactification of $\mathbb{R}^3$ with the point of the infinity) in the canonical way. Let T_2 be another solid torus embedded inside T_1 as shown in figure 8 and $h : \mathbb{S}^3 \to \mathbb{S}^3$ the homeomorphism of $\mathbb{S}^3$ sending T_1 to $T_2 = h(T_1)$. This is possible since they are tubular neighborhoods of two equivalent (trivial) knots. Consider the iterations of T_1 by h,

$$T_2 = h(T_1), T_3 = h^2(T_1), \dots, T_n = h^{n-1}(T_1), \dots$$

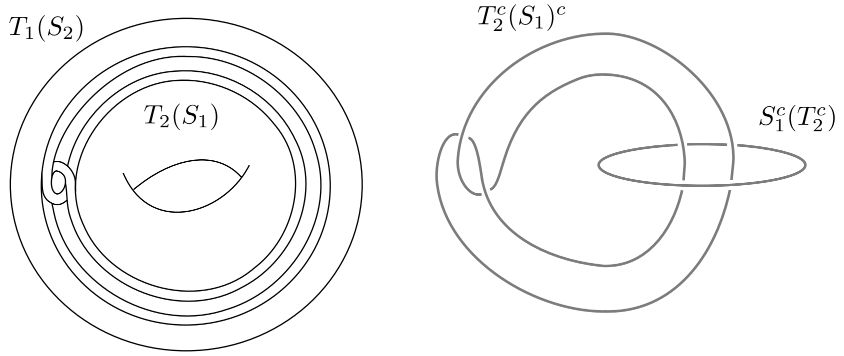


Figure 8: The first stage of the construction of the Whitehead continuum and the Whitehead link. The elements within the parentheses represent the effect of performing the interchange of the two components of the Whitehead link.

to obtain an infinite sequence of nested solid tori,

$$T_1 \supset T_2 \supset \ldots \supset T_n \supset T_{n+1} \supset \ldots$$

where, for every $n \in \mathbb{N}$, T_{n+1} is embedded in T_n as T_2 is in T_1. Let us consider the intersection of all the tori

$$wh = \bigcap_{n \in \mathbb{N}} T_n$$

which, as an intersection of connected, closed and nested sets, is a compact connected space, called the Whitehead continuum. Then, considering the complementary in $\mathbb{S}^3$ of this continuum, we define the Whitehead manifold $\mathcal{W} = \mathbb{S}^3 \setminus wh$. This is the contractible open connected 3-manifold introduced by Whitehead in [17] as a counterexample to a false theorem of him self concerning the Poincaré conjecture that he claimed in [16].

Now, let us come back to dimension 4. It turns out that the cartesian product of the Whitehead manifold and the real line is homeomorphic to $\mathbb{R}^4$, even being $\mathcal{W}$ different to $\mathbb{R}^3$. Namely,

Theorem 5.1 *Let $\mathcal{W}$ be the Whitehead manifold. Then the space $\mathcal{W} \times \mathbb{R}$ is homeomorphic to $\mathbb{R}^4$.*

This was first observed by Shapiro, probably in private conversations with several authors (Bing among others) and this is the source of manifold factors theories. Indeed, Bing, in [6], mentions that it was thinking about the result of Shapiro when he realized that he could find 4-cells separating different levels of solid torus cross closed intervals in the definition of the Whitehead manifold just playing with the temporal coordinate, as done in theorem 4.2. Shapiro's result was later proved by Glimm [12] where he also shows that $\mathcal{W} \times \mathcal{W}$ is homeomorphic to $\mathbb{R}^6$. McMillan [13] proved the result for a kind of manifold that generalizes $\mathcal{W}$ and also showed that every contractible manifold cross $\mathbb{R}^n$ is homeomorphic to $\mathbb{R}^{2n}$.

We give a complete proof here of Shapiro's result, suggested by Bing in [6] using a version of theorem 4.2 adapted to the Whitehead manifold and the following theorem of Morton Brown.

Theorem 5.2 (M.Brown [10]) *A topological space which is the union of an increasing sequence of open sets*

$$U_1 \subset U_2 \subset \ldots \subset U_n \subset \ldots,$$

such that all of them are homeomorphic to $\mathbb{R}^n$*, is homeomorphic to* $\mathbb{R}^n$*.*

Now we prove the previous theorem. The idea is to write the Whitehead manifold as a union of solid tori and place 4-cells between the terms of the sequence formed by the cartesian products in order to apply the theorem of Morton Brown.

Proof. (Theorem 5.1) Using the De Morgan laws and the fact that the solid tori in the definition are nested, we can write the Whitehead manifold as a union of sets, as follows:

$$\mathcal{W} = \mathbb{S}^3 \setminus \bigcap_{n\in\mathbb{N}} T_n = \bigcup_{n\in\mathbb{N}} \mathbb{S}^3 \setminus T_n = \bigcup_{n\in\mathbb{N}} \mathbb{S}^3 \setminus \operatorname{Int} T_n.$$

In fact, each of the $S_n = \mathbb{S}^3 \setminus \operatorname{Int} T_n$ is an unkotted solid tori, since each T_n is. Moreover, we claim that each S_n is embedded in S_{n+1} exactly in the same way T_{n+1} is embedded in T_n. This is just an exercise of visualization: Consider the link in figure 8, called, for obvious reasons, the Whitehead link. One component represents the *core* (the curve where the solid torus could be retracted onto, written T^c for a solid torus T, in the figure) of S_1 and the other one the core of T_2. But it turns out that the Whitehead link is interchangeable, that is, there exists a homeomorphism g of $\mathbb{S}^3$ that changes one componet of the link into the other (the best way to visualize this is to produce the link physically and convince our selves doing the interchange). Hence, if that link represents cores of solid tori, the homeomorphism g of $\mathbb{S}^3$ exchanges the position of the corresponding solid tori. It is possible to proceed the same way for the rest of indices. So we can write the Whitehead manifold as an ascending union of solid tori $\mathcal{W} = \bigcup_{n\in\mathbb{N}} S_n$ such that S_n is embedded in S_{n+1} as T_{n+1} is in T_n. That is, as T_2 is in T_1 in figure 8.

Observe that we have the homeomorphism

$$\mathcal{W} \times \mathbb{R} = \bigcup_{n\in\mathbb{N}} S_n \times \mathbb{R} = \bigcup_{n\in\mathbb{N}} S_n \times [-n, n],$$

because $S_n \subset S_{n+1}$ for every $n \in \mathbb{N}$.

Finally, we will find useful 4-cells in Bing's style in order to apply Morton Brown's theorem. We proceed as in theorem 4.2, but only with the left part and less number of tori. Consider the first step in the constrcution of the Whitehead manifold, that is, S_1 embedded inside S_2, as shown in figure 8. Let $[a, b]$ be any closed interval and $\varepsilon > 0$ a real number. Cut the solid torus S_2 with two disks not disturbing the knotted part of S_1 as we did in the Dogbone construction and we get two 3-cells, C_L and C_M with the same notation as there. Consider also in the same way C_L^+ and C_L^- removing the corresponding thickenings of the disks. Then we claim that the union of the five 4-cells

$$\begin{aligned}
&C_B = C_M \times [a-\varepsilon, b+\varepsilon], \\
&C_{LW}^+ = C_L^+ \times [a-\varepsilon, a-\frac{\varepsilon}{2}], \qquad C_{LW}^- = C_L^- \times [b+\frac{\varepsilon}{2}, b+\varepsilon], \\
&C_{L1} = (C_L \cap S_1) \times [a-\frac{\varepsilon}{2}, b], \quad C_{L2} = (C_L \cap S_1) \times [a, b+\frac{\varepsilon}{2}],
\end{aligned}$$

forms a 4-cell K such that

$$S_1 \times [a, b] \subset \operatorname{Int} K \subset K \subset S_2 \times [a-\varepsilon, b+\varepsilon].$$

We can find 4-cells in every level of the construction in the same way we did in corollary 7. So, for every $n \in \mathbb{N}$, we can find a 4-cell K_n such that

$$S_n \times [-n, n] \subset \operatorname{Int} K_n \subset K_n \subset S_{n+1} \times [-(n+1), n+1].$$

Hence, we can write

$$\mathcal{W} \times \mathbb{R} = \bigcup_{n \in \mathbb{N}} \operatorname{Int} K_n.$$

By theorem 5.2, $\mathcal{W} \times \mathbb{R}$ is homeomorphic to $\mathbb{R}^4$. □

References

[1] J. W. Alexander: *An Example of a Simply Connected Surface Bounding a Region which is not Simply Connected*. Proc. Nat. Acad. Sci. U.S.A. **10** (1924) 8–10.

[2] J. J. Andrews, M. L. Curtis: *n-Space modulo and arc*. Ann. of Math. **75** (1962) 1–7.

[3] S. Armentrout: *Monotone decompositions of* E^3. In *Topology seminar (Wisconsin, 1965),*Ann. of Math. Studies **60** (1966) 1–25.

[4] R. H. Bing: *A homeomorphism between the* 3*-sphere and he sum of two solid horned spheres*. Ann. of Math. **56** (1952) 354–362.

[5] R. H. Bing: *A decomposition of* E^3 *into points and tame arcs such that the decomposition space is topologically different from* E^3. Ann. of Math. **65** (1957) 484–500.

[6] R. H. Bing: *The cartesian product of a certain nonmanifold and a line is* E^4. Ann. of Math. **70** (1959) 399–412.

[7] R. H. Bing: *A set is a* 3 *cell if its cartesian product with an arc is a* 4 *cell*. Proc. AMS **12** (1961) 13–19.

[8] R. J. Daverman: *Decompositions of manifolds*. Reprint of the 1986 original. AMS Chelsea Publishing. (2007), xii+317 pp.

[9] M. Brown: *A proof of the generalized Schoenflies theorem*. Bull. AMS **66** (1960) 74–76.

[10] M. Brown: *The monotone union of open n-cells is an open n-cell*. Notices AMS **12** (1961) 812–814.

[11] R. H. Fox, E. Artin: *Some wild cells and spheres in three-dimensional space*. Ann. of Math. **49** (1948) 979–990.

[12] J. Glimm: *Two cartesian products which are euclidean spaces*. Bull. Soc. Math. France **88** (1960) 131–135.

[13] D. R. McMillan Jr.: *Cartesian products of contractible open manifolds*. Bull. AMS **67** (1961), 510–514.

[14] R. L. Moore: *Concerning upper semicontinuous collections of continua*. Trans. AMS **27** (1925) 416–428.

[15] L. R. Rubin: *A general class of factors of E^4*. Trans. AMS **166** (1972), 215-224.

[16] J. H. C. Whitehead: *Certain theorems about three-dimensional manifolds (I)*. Quart. J. of Math. (Oxford) **5** (1934) 308–320.

[17] J. H. C. Whitehead: *A certain open manifold whose group is unity*. Quart. J. of Math. (Oxford) **6** (1935) 268–279.

[18] G. T. Whyburn: *On the structure of continua*. Bull. AMS **42** (1936) 49–73.

Equivariant motive of the $\mathrm{SL}(3,\mathbb{C})$-character variety of torus knots

Vicente Muñoz and Jonathan Sánchez

Facultad de Matemáticas,
Universidad Complutense de Madrid,
Plaza Ciencias 3, 28040 Madrid, Spain
vicente.munoz@mat.ucm.es jnsanchez@mat.ucm.es

Dedicated to José María Montesinos Amilibia, with our deepest admiration.

ABSTRACT

Let Γ be the fundamental group of the complement of the torus knot of type (m,n). This has a presentation $\Gamma = \langle x, y \,|\, x^m = y^n \rangle$. Using the geometric description of the character variety $X(\Gamma, G)$ of characters of representations of Γ into $G = \mathrm{SL}(3,\mathbb{C})$, we determine explicitly its associated μ_3-equivariant motive.

1. Introduction

Let Γ be a finitely presented group, and let $G = \mathrm{SL}(r,\mathbb{C})$. A *representation* of Γ in G is a homomorphism $\rho : \Gamma \to G$. Consider a presentation $\Gamma = \langle x_1, \ldots, x_k | r_1, \ldots, r_s \rangle$. Then ρ is completely determined by the k-tuple $(A_1, \ldots, A_k) = (\rho(x_1), \ldots, \rho(x_k))$ subject to the relations $r_j(A_1, \ldots, A_k) = \mathrm{Id}$, $1 \le j \le s$. The space of representations is

$$\begin{aligned} R(\Gamma, G) &= \mathrm{Hom}(\Gamma, G) \\ &= \{(A_1, \ldots, A_k) \in G^k \,|\, r_j(A_1, \ldots, A_k) = \mathrm{Id},\, 1 \le j \le s\} \subset G^k \,. \end{aligned}$$

Therefore $R(\Gamma, G)$ is an affine algebraic set.

We say that two representations ρ and ρ' are equivalent if there exists $P \in G$ such that $\rho'(g) = P^{-1}\rho(g)P$, for every $g \in G$. The moduli space of representations is defined as the GIT quotient

$$M(\Gamma, G) = R(\Gamma, G) // G \,.$$

Recall that by definition of GIT quotient for an affine variety, if we write $R(\Gamma, G) = \mathrm{Spec}\, A$, then $M(\Gamma, G) = \mathrm{Spec}\, A^G$. For a representation $\rho : \Gamma \to G$, we define its *character* as the map $\chi_\rho : \Gamma \to \mathbb{C}$, $\chi_\rho(g) = \mathrm{tr}\,\rho(g)$. Note that two equivalent representations ρ and ρ' have the same character. There is a character map $\chi : R(\Gamma, G) \to \mathbb{C}^\Gamma$, $\rho \mapsto \chi_\rho$, whose image

$$X(\Gamma, G) = \chi(R(\Gamma, G))$$

is called the *character variety of* Γ. The traces χ_ρ span a subring $B \subset A^G$, and $X(\Gamma, G) = \operatorname{Spec} B$. Actually, for $G = \mathrm{SL}(r, \mathbb{C})$, the ring of invariant polynomials is generated by characters (see Chapter 1 in [12]), so the natural algebraic map

$$M(\Gamma, G) \to X(\Gamma, G)$$

is an isomorphism.

The character varieties for $\mathrm{SL}(2, \mathbb{C})$ have been extensively studied in the last three decades [3, 4, 12]. Given a manifold M, the moduli of representations of $\pi_1(M)$ into $\mathrm{SL}(2, \mathbb{C})$ contain information of the topology of M. This is specially relevant for 3-dimensional manifolds [3], where the fundamental group and the geometrical properties of the manifold are strongly related. This has been used to study knots $K \subset S^3$, by analysing the $\mathrm{SL}(2, \mathbb{C})$-character variety of the fundamental group of the knot complement $S^3 - K$ (these are called *knot groups*). The case of $\mathrm{SL}(2, \mathbb{C})$-representations of the fundamental group of a surface has also been extensively analysed [5, 7, 11, 15], in this situation focusing more on geometrical properties of the moduli space in itself (cf. non-abelian Hodge theory).

However, much less is known of the character varieties for other groups, notably for $\mathrm{SL}(r, \mathbb{C})$ with $r \geq 3$. The character varieties for $\mathrm{SL}(3, \mathbb{C})$ for free groups have been described in [9, 10]. In the case of 3-manifolds, little has been done. For knot groups, the first case to analyse is clearly that of torus knots. These are defined as follows. Let $T^2 = S^1 \times S^1$ be the 2-torus and consider the standard embedding $T^2 \subset S^3$. Let m, n be a pair of coprime positive integers. Identifying T^2 with the quotient $\mathbb{R}^2/\mathbb{Z}^2$, the image of the straight line $y = \frac{m}{n}x$ in T^2 defines the *torus knot* of type (m, n), which we shall denote as $K_{m,n} \subset S^3$ (see [18, Chapter 3]). The $\mathrm{SL}(3, \mathbb{C})$-character variety of the torus knot $K_{2,3}$ has been described in [6], and for the general torus knot $K_{m.n}$ it is given in [17].

The fundamental group of the knot complement $S^3 - K_{m,n}$ is the group

$$\Gamma_{m,n} = \langle x, y \,|\, x^n = y^m \rangle \,.$$

Therefore the character variety is described explicitly as

$$\mathcal{X}_r = X(\Gamma_{m,n}, \mathrm{SL}(r, \mathbb{C})) = \{(A, B) \in \mathrm{SL}(r, \mathbb{C})^2 \,|\, A^n = B^m\} // \mathrm{SL}(r, \mathbb{C}) \tag{1.1}$$

Various geometrical properties of character varieties can be studied. Basic properties include connectedness, number of irreducible components, and the dimension. More elaborated properties are the fundamental group or the Poincaré polynomials; such topological properties have been studied for the character varieties for surfaces via non-abelian Hodge theory, which produces a *homeomorphism* of the moduli of representations with the moduli space of Higgs bundles [7]. If one focuses on the algebro-geometric aspects of character varieties, one can try to compute the motives, the Hodge numbers or the E-polynomials. For instance, for $\mathrm{SL}(3, \mathbb{C})$-character varieties of torus knots, the motive is given in [17].

Here, we shall give, using the result of [17], the μ_3-equivariant motive of the $\mathrm{SL}(3, \mathbb{C})$-character varieties of torus knots. Note that the center of $\mathrm{SL}(r, \mathbb{C})$, consisting of the matrices $\varpi\,\mathrm{Id}$, where $\varpi \in \mu_r = \{e^{2\pi ik/r}, k = 0, \ldots, r-1\}$, act on (1.1). Therefore the motive of $X(\Gamma_{m,n}, \mathrm{SL}(r,\mathbb{C}))$ has a μ_r-action. This produces a μ_r-equivariant motive as explained in Section 2. Our main result is:

Theorem 1.1 *The μ_3-equivariant motive of the* $\mathrm{SL}(3,\mathbb{C})$*-character variety of the* (m,n)*-torus knot is:*

- *If $n, m \equiv 1, 5 \pmod 6$, then*

$$\begin{aligned} h_{\mu_3}(\mathcal{X}_3) = \Big[&P_0 + \frac{1}{36}(m-1)(m-2)(n-1)(n-2)P_1 + \frac{1}{6}(n-1)(m-1)(n+m-4)P_3 + \\ &+ \frac{1}{4}(n-1)(m-1)P_5\Big]T + \\ &+ \Big[\frac{1}{3}(m-1)(n-1)(n+m-4)P_3 + \frac{1}{18}(m-2)(m-1)(n-2)(n-1)P_1\Big]R \end{aligned}$$

- *If $n \equiv 2, 4 \pmod 6$, $m \equiv 1, 5 \pmod 6$, then*

$$\begin{aligned} h_{\mu_3}(\mathcal{X}_3) = \Big[&P_0 + \frac{1}{36}(m-1)(m-2)(n-1)(n-2)P_1 + \frac{1}{6}(n-1)(m-1)(n+m-4)P_3 + \\ &+ \frac{1}{4}(n-2)(m-1)P_5 + \frac{1}{2}(m-1)P_6\Big]T + \\ &+ \Big[\frac{1}{3}(m-1)(n-1)(n+m-4)P_3 + \frac{1}{18}(m-2)(m-1)(n-2)(n-1)P_1\Big]R \end{aligned}$$

- *If $n \equiv 3 \pmod 6$, $m \equiv 1, 5 \pmod 6$, then*

$$\begin{aligned} h_{\mu_3}(\mathcal{X}_3) = \Big[&P_0 + \frac{1}{36}(m-1)(m-2)n(n-3)P_1 + \frac{1}{6}(m-1)(m-2)P_2 + \\ &+ \frac{1}{6}(m-1)(mn+n^2-5n-m-2)P_3 + (m-1)P_4 + \frac{1}{4}(n-1)(m-1)P_5\Big]T + \\ &+ \Big[\frac{1}{18}(m-2)(m-1)(n^2-3n+3)P_1 - \frac{1}{6}(m-2)(m-1)P_2 + \\ &+ \frac{1}{3}(m-1)(n^2+mn-5n-m+7)P_3 - (m-1)P_4\Big]R \end{aligned}$$

- *If $n \equiv 0 \pmod 6$, $m \equiv 1, 5 \pmod 6$, then*

$$\begin{aligned} h_{\mu_3}(\mathcal{X}_3) = \Big[&P_0 + \frac{1}{36}(m-1)(m-2)n(n-3)P_1 + \frac{1}{6}(m-1)(m-2)P_2 + \\ &+ \frac{1}{6}(m-1)(mn+n^2-5n-m-2)P_3 + (m-1)P_4 + \\ &+ \frac{1}{4}(n-2)(m-1)P_5 + \frac{1}{2}(m-1)P_6\Big]T + \\ &+ \Big[\frac{1}{18}(m-2)(m-1)(n^2-3n+3)P_1 - \frac{1}{6}(m-2)(m-1)P_2 + \\ &+ \frac{1}{3}(m-1)(n^2+mn-5n-m+7)P_3 - (m-1)P_4\Big]R \end{aligned}$$

- *If* $n \equiv 2,4 \pmod 6$, $m \equiv 3 \pmod 6$, *then*

$$\begin{aligned} h_{\mu_3}(\mathcal{X}_3) = \Big[& P_0 + \frac{1}{36} m(m-3)(n-1)(n-2)P_1 + \frac{1}{6}(n-1)(n-2)P_2 + \\ & + \frac{1}{6}(n-1)(mn+m^2-n-5m-2)P_3 + (n-1)P_4 + \\ & + \frac{1}{4}(n-2)(m-1)P_5 + \frac{1}{2}(m-1)P_6 \Big] T + \\ & + \Big[\frac{1}{18}(n-2)(n-1)(m^2-3m+3)P_1 - \frac{1}{6}(n-2)(n-1)P_2 + \\ & + \frac{1}{3}(n-1)(m^2+mn-5m-n+7)P_3 - (n-1)P_4 \Big] R \end{aligned}$$

where T is the trivial representation and R is the non-trivial two-dimensional rational representation. Here, $P_0 = \mathbb{L}^2$, $P_1 = \mathbb{L}^4 + 4\mathbb{L}^3 - 3\mathbb{L}^2 - 15\mathbb{L} + 12$, $P_2 = \mathbb{L}^4 + 2\mathbb{L}^3 - 3\mathbb{L}^2 - \mathbb{L} + 4$, $P_3 = \mathbb{L}^2 - 3\mathbb{L} + 3$, $P_4 = \mathbb{L}^2 - \mathbb{L} + 1$, $P_5 = \mathbb{L}^2 - 3\mathbb{L} + 2$, $P_6 = \mathbb{L}^2 - 2\mathbb{L} + 1$.

(Note that we can swap n, m if necessary to be in one of the cases above.)

2. Equivariant motives

Let $\mathcal{V}ar_{\mathbb{C}}$ be the category of quasi-projective complex varieties. We denote by $K(\mathcal{V}ar_{\mathbb{C}})$ the Grothendieck ring of $\mathcal{V}ar_{\mathbb{C}}$. This is the abelian group generated by elements $[Z]$, for $Z \in \mathcal{V}ar_{\mathbb{C}}$, subject to the relation $[Z] = [Z_1] + [Z_2]$ whenever Z can be decomposed as a disjoint union $Z = Z_1 \sqcup Z_2$ of a closed and a Zariski open subset. There is a naturally defined product in $K(\mathcal{V}ar_{\mathbb{C}})$ given by $[Y] \cdot [Z] = [Y \times Z]$. We write $\mathbb{L} := [\mathbb{A}^1]$, where $\mathbb{A}^1$ is the affine line, the *Lefschetz object* in $K(\mathcal{V}ar_{\mathbb{C}})$. Clearly $\mathbb{L}^k = [\mathbb{A}^k]$. Finally, let $\mathcal{S}m\mathcal{V}ar_{\mathbb{C}}$ denote the category of *smooth projective* varieties over $\mathbb{C}$. We consider the ring $K^{bl}(\mathcal{S}m\mathcal{V}ar_{\mathbb{C}})$ generated by the smooth projective varieties subject to the relations $[X] - [Y] = [\mathrm{Bl}_Y(X)] - [E]$, where $Y \subset X$ is a smooth subvariety, $\mathrm{Bl}_Y(X)$ is the blow-up of X along Y, and E is the exceptional divisor. By [2, Theorem 3.1], there is an isomorphism

$$K^{bl}(\mathcal{S}m\mathcal{V}ar_{\mathbb{C}}) \cong K(\mathcal{V}ar_{\mathbb{C}}).$$

Now we move to the definition of Chow motives. Given a smooth projective variety X, let $CH^d(X)$ denote the abelian group of $\mathbb{Q}$-cycles on X, of codimension d, modulo rational equivalence. If $X, Y \in \mathcal{S}m\mathcal{V}ar_{\mathbb{C}}$, suppose that X is connected and $\dim(X) = d$. The group of correspondences (of degree 0) from X to Y is $\mathrm{Corr}(X,Y) = CH^d(X \times Y)$. For varieties $X, Y, Z \in \mathcal{S}m\mathcal{V}ar_{\mathbb{C}}$, the composition of correspondences

$$\mathrm{Corr}(X,Y) \otimes \mathrm{Corr}(Y,Z) \to \mathrm{Corr}(X,Z)$$

is defined as

$$g \circ f = p_{XZ*}(p^*_{XY}(f) \cdot p^*_{YZ}(g)),$$

where $p_{XZ} : X \times Y \times Z \to X \times Z$ is the projection, and similarly for p_{XY} and p_{YZ}.

Definition 2.1 *The category of* (effective Chow) motives *is the category $\mathcal{M}ot$ such that:*

- *its objects are pairs (X,p) where $X \in \mathcal{S}m\mathcal{V}ar_{\mathbb{C}}$, and $p \in \mathrm{Corr}(X,X)$ is an idempotent ($p = p \circ p$);*
- *if $(X,p),(Y,q)$ are effective motives, then the morphisms are* $\mathrm{Hom}((X,p),(Y,q)) = q \circ \mathrm{Corr}(X,Y) \circ p$.

There is a natural functor

$$h : \mathcal{S}m\mathcal{V}ar_{\mathbb{C}}^{\mathrm{opp}} \to \mathcal{M}ot \tag{2.1}$$

such that, for a smooth projective variety X, $h(X) = (X, \Delta_X)$, where $\Delta_X \in \mathrm{Corr}(X,X)$ is the graph of the identity $\mathrm{Id}_X : X \to X$. We say that $h(X)$ is the *motive of* X.

The category $\mathcal{M}ot$ is pseudo-abelian, where direct sums and tensor products are defined by $(X,p)\oplus(Y,q) = (X\sqcup Y, p+q)$ and $(X,p)\otimes(Y,q) = (X\times Y, p^*_{X\times X}(p)\cdot p^*_{Y\times Y}(q))$. In particular

$$\begin{aligned} h(X \sqcup Y) &= h(X) \oplus h(Y), \\ h(X \times Y) &= h(X) \otimes h(Y). \end{aligned}$$

This allows us to define $K(\mathcal{M}ot)$ as the abelian group generated by elements $[M]$, for $M \in \mathcal{M}ot$, subject to the relations $[M] = [M_1] + [M_2]$, when $M = M_1 \oplus M_2$. This is a ring with the product $[M_1]\cdot[M_2] = [M_1 \otimes M_2]$.

In $\mathcal{M}ot$, we have that $\mathbf{1} = h(pt)$ is the identity of the tensor product, so it is called the *unit motive*. It is easily seen that there is an isomorphism $\mathbf{1} \cong (\mathbb{P}^1, \mathbb{P}^1 \times pt)$. Set $\mathbb{L} = (\mathbb{P}^1, pt \times \mathbb{P}^1)$, which is called the *Lefschetz motive*. Therefore $h(\mathbb{P}^1) = \mathbf{1} \oplus \mathbb{L}$, and more generally,

$$h(\mathbb{P}^n) = \mathbf{1} \oplus \mathbb{L} \oplus \cdots \oplus \mathbb{L}^n.$$

Denote also by $\mathbb{L} \in K(\mathcal{M}ot)$ the class of the Lefschetz motive $\mathbb{L} \in \mathcal{M}ot$.

In [13] it is shown that the motive of the blow-up of a smooth projective variety X along a codimension r smooth subvariety Y is $h(\mathrm{Bl}_Y(X)) = h(X) \oplus \left(\bigoplus_{i=1}^{r-1}(h(Y)\otimes\mathbb{L}^i)\right)$, being thus compatible with the relation defining $K^{bl}(\mathcal{S}m\mathcal{V}ar_{\mathbb{C}})$. So the map h in (2.1) descends to $K^{bl}(\mathcal{S}m\mathcal{V}ar_{\mathbb{C}}) \to K(\mathcal{M}ot)$, hence defining a ring homomorphism

$$\chi : K(\mathcal{V}ar_{\mathbb{C}}) \to K(\mathcal{M}ot). \tag{2.2}$$

When X is smooth and projective, we have

$$\chi([X]) = [h(X)],$$

so we can think of the map χ as the natural extension of the notion of motives to all quasi-projective varieties. Notice that $\chi(\mathbb{L}) = \mathbb{L}$, which justifies the use of the same notation for the Lefschetz object and the Lefschetz motive.

Let G be a finite group. We have the category $\mathcal{V}ar_{\mathbb{C}}^G$ of quasi-projective complex varieties with a G-action, and the category $\mathcal{S}m\mathcal{V}ar_{\mathbb{C}}^G$ of smooth projective complex varieties endowed with a G-action. As before, we have well-defined Grothendieck rings $K(\mathcal{V}ar_{\mathbb{C}}^G)$ and $K^{bl}(\mathcal{S}m\mathcal{V}ar_{\mathbb{C}}^G)$, which are isomorphic [2].

Let X be a smooth projective variety with an action of a finite group G. Let $CH^d_{\mathbb{C}}(X\times X) = CH^d(X \times X) \otimes \mathbb{C}$ denote the Chow ring with complex coefficients. The action of G on X defines a morphism

$$\varphi : \mathbb{C}[G] \longrightarrow CH^d_{\mathbb{C}}(X \times X),$$

given by $g \mapsto \Gamma_g$. By a theorem of Maschke [8, XVIII, Thm, 1.2], the group ring $\mathbb{C}[G]$ is semisimple. Every semisimple ring R admit a decomposition in simple rings $R = \prod_{i=1}^s R_i$, where $R_i = R \cdot e_i$. Such $e_i \in R$ are the idempotents of R_i and $e_i \cdot e_j = 0$ for $i \neq j$. Furthermore, the sum of these elements is

$$1 = e_1 + e_2 + \cdots + e_s. \tag{2.3}$$

In our case,

$$\mathbb{C}[G] = \prod_{i=1}^{s} \mathbb{C}[G] \cdot e_i \tag{2.4}$$

where $e_i^2 = e_i$ and $e_i \cdot e_j = 0$ whenever $i \neq j$. If we let $p_i = \varphi(e_i) \in CH^d_{\mathbb{C}}(X \times X)$, then $p_i^2 = p_i$ and $p_i \cdot p_j = 0$ for $i \neq j$. The equality (2.3) gives the decomposition of the motive $h(X)$ of the variety

$$h(X) = \bigoplus_{i=1}^{s} (X, p_i).$$

Definition 2.2 *We define the* equivariant motive *of X as*

$$h_G(X) := \sum (X, p_i) e_i \in K(\mathcal{M}ot) \otimes \mathbb{C}[G].$$

This means that $h_G(X)$ is the image of $\sum e_i \otimes e_i \in \mathbb{C}[G] \otimes \mathbb{C}[G]$ under the natural map $\varphi \otimes \mathrm{Id} : \mathbb{C}[G] \otimes \mathbb{C}[G] \to CH^d_{\mathbb{C}}(X \times X) \otimes \mathbb{C}[G]$.

The proof of [13] can be carried out for a smooth projective variety X endowed with a G-action and a smooth subvariety $Y \subset X$ which is G-invariant. This gives that $h_G(\mathrm{Bl}_Y(X)) = h_G(X) \oplus \left(\bigoplus_{i=1}^{r-1} (h_G(Y) \otimes \mathbb{L}^i)\right)$. Thus the map h_G in Definition 2.2 descends to a map

$$K(\mathcal{V}ar^G_{\mathbb{C}}) = K^{bl}(\mathcal{S}m\mathcal{V}ar^G_{\mathbb{C}}) \to K(\mathcal{M}ot) \otimes \mathbb{C}[G].$$

The proof of this fact follows the same arguments presented in [13], taking into account the equivariance of the Chern classes x_k of the projective bundle $E \to Y$, where E denotes the exceptional divisor of $\mathrm{Bl}_Y X$, and hence the classes p_i commute with x_k.

The idempotents e_i are associated in a one-to-one way to the irreducible representations of G. For an irreducible representation R_i, let χ_i be its character. Then

$$e_i = \frac{1}{|G|} \sum_{g \in G} \chi_i(g) g,$$

so $h_i(X) = (X, p_i)$, where

$$p_i = \frac{1}{|G|} \sum_{g \in G} \chi_i(g) \Gamma_g.$$

For the trivial representation R_1, we recover the quotient motive of X/G by the result in [1],

$$h(X/G) = h_1(X) = \left(X, \frac{1}{|G|} \sum_{g\in G} \Gamma_g \right). \tag{2.5}$$

This holds for smooth projective varieties by [1], hence it holds for all quasi-projective varieties since $K(\mathcal{V}ar_{\mathbb{C}}^G) = K^{bl}(\mathcal{S}m\mathcal{V}ar_{\mathbb{C}}^G)$.

Finally, from $h(X) = \sum h_i(X)$ the equivariant motive recovers the usual motive of a quasi-projective variety.

Now we analyse the case of a cyclic group $G = C_r$ of order r.

Lemma 2.1 *Let ξ be an r-th primitive root of unity and let g be a generator for the group C_r. Then, the decomposition* (2.4) *is*

$$\mathbb{C}[C_r] = \bigoplus_{a=0}^{r-1} \mathbb{C}\, e_a, \quad \textit{where the projectors are } \; e_a = \frac{1}{r}\sum_{k=0}^{r-1} (\xi^a g)^k .$$

Proof. First, we compute the product $e_a \cdot e_b$. By definition

$$\begin{aligned} e_a \cdot e_b &= \frac{1}{r^2} \left(\sum_{i=0}^{r-1} (\xi^a g)^i \right) \left(\sum_{j=0}^{r-1} (\xi^b g)^j \right) \\ &= \frac{1}{r^2} \sum_{c=0}^{r-1} \sum_{\substack{i+j\equiv c \\ (\mathrm{mod}\ r)}} \xi^{ai+bj} g^{i+j} = \frac{1}{r^2} \sum_{c=0}^{r-1} g^c \sum_{\substack{i+j\equiv c \\ (\mathrm{mod}\ r)}} \xi^{ai+bj}. \end{aligned}$$

We focus on the sum $\sum_{i+j=c} \xi^{ai+bj}$. If $a \neq b$, this sum is zero, since the sequence $\{ai+bj \pmod r\}_{i+j=c}$ is nothing but $\{0,1,\ldots,r-1\}$. Thus, $e_a \cdot e_b = 0$ if $a \neq b$. The case $a = b$, the sum is non-zero and it is $r \cdot (\xi^a)^c$, and we conclude that $e_a \cdot e_a = e_a$. □

In Lemma 2.1, the element corresponding to the trivial representation is $e_0 = \frac{1}{r}\sum g^k$. So $h_0(X) = h(X/C_r)$. Suppose that we are in the situation

$$h_a(X) = h_b(X), \text{ when } \gcd(r,a) = \gcd(r,b). \tag{2.6}$$

Then we can recover the equivariant motive from the quotients $X/\langle g^d \rangle$, for $d|r$. We start with the case that r is prime. Then $h_1(X) = \ldots = h_{r-1}(X)$. Hence

$$h_{C_r}(X) = h_0(X) e_0 + h_1(X)(e_1 + \ldots + e_{r-1}), \tag{2.7}$$

where

$$\begin{aligned} h_0(X) &= h(X/C_r), \\ h_1(X) &= \frac{1}{r-1}(h(X) - h(X/C_r)). \end{aligned}$$

If r is not prime, then

$$h_{C_r}(X) = h_0(X)e_0 + \sum_{d|r} h_d(X) \left(\sum_{\substack{1\le l\le r/d-1 \\ \gcd(l,r/d)=1}} e_{ld} \right).$$

To determine $h_{C_r}(X)$ we need as many equations as divisors of r. These are provided by the following result.

Lemma 2.2 *For any $d|r$, we have $\sum_{k=0}^{r/d-1} h_{kd}(X) = h(X/\langle g^{r/d}\rangle)$.*

Proof. Let ξ be an r-th primitive root of the unity and let $g \in G$ be a generator of the group C_r. Then,

$$\sum_{k=0}^{r/d-1} e_{kd} = \frac{1}{r}\sum_{k=0}^{r/d-1}\sum_{i=0}^{r-1}\left(\xi^{kd}g\right)^i = \frac{1}{r}\sum_{k=0}^{r/d-1}\sum_{i=0}^{r-1}\xi^{kdi}g^i = \frac{1}{r}\sum_{i=0}^{r-1} g^i \sum_{k=0}^{r/d-1}\xi^{kdi}.$$

The sum $\sum_{k=0}^{r/d-1}\xi^{kdi}$ is zero if and only if di is multiple of r, that is, $i = \frac{r}{d}b$ for some integer number b. Then, the sum becomes

$$\sum_{k=0}^{r/d-1} e_{kd} = \frac{1}{d}\sum_{b=0}^{d-1} g^{\frac{r}{d}b}.$$

Now take the image under $\varphi : \mathbb{C}[C_r] \to CH^d_{\mathbb{C}}(X\times X)$. This produces the motive

$$\left(X, \frac{1}{d}\sum_{b=0}^{d-1}\Gamma_{g^{\frac{r}{d}b}}\right) = h(X/\langle g^{r/d}\rangle).$$

The result follows. □

3. Character varieties of torus knots

Let

$$\Gamma_{m,n} = \langle x, y \,|\, x^n = y^m \rangle$$

be the torus knot group, and consider the character varieties for $\mathrm{SL}(r,\mathbb{C})$ and $\mathrm{PGL}(r,\mathbb{C}) = \mathrm{SL}(r,\mathbb{C})/\mu_r$,

$$\begin{aligned} \mathcal{X}_r &= X(\Gamma_{m,n}, \mathrm{SL}(r,\mathbb{C})), \\ \bar{\mathcal{X}}_r &= X(\Gamma_{m,n}, \mathrm{PGL}(r,\mathbb{C})). \end{aligned}$$

By [17, Section 4], we have that μ_r acts on $\mathcal{X}_r$ via $\varpi\cdot(A,B) = (\varpi^m A, \varpi^n B)$, $\varpi\in\mu_r$, and

$$\bar{\mathcal{X}}_r \cong \mathcal{X}_r/\mu_r\,.$$

Let us see that Condition (2.6) is satisfied for this action. Take a,b such that $\gcd(a,r) = \gcd(b,r) = d$. Then there is a Galois automorphism $\sigma : \mathbb{C} \to \mathbb{C}$ such that $\sigma(\xi^b) = \xi^a$, where $\xi = e^{2\pi i/r}$. We let σ act on $\mathcal{X}_r$: this means that σ acts on all entries of the matrices A and B. Then $\sigma : \mathcal{X}_r \to \mathcal{X}_r$ interchanges the action of g to the action of $\sigma(g) = g^p$, where p is defined by $\xi^{bp} = \xi^a$, i.e. $\frac{b}{d}p \equiv \frac{a}{d} \pmod{\frac{r}{d}}$ (the integer p is coprime to r). Therefore $(\mathcal{X}_r, p_a) \cong (\mathcal{X}_r, \sigma(p_a)) = (\mathcal{X}_r, p_b)$, since

$$\sigma(p_a) = \frac{1}{r}\sum_{k=0}^{r-1}(\xi^a\sigma(g))^k = \frac{1}{r}\sum_{k=0}^{r-1}(\xi^a g^p)^k = \frac{1}{r}\sum_{k=0}^{r-1}\left(\xi^{bp}g^p\right)^k = \frac{1}{r}\sum_{k=0}^{r-1}(\xi^b g)^k = p_b.$$

We have the following result for $\mathrm{SL}(2,\mathbb{C})$-character varieties.

Theorem 3.1 *The μ_2-equivariant motive $h_{\mu_2}(\mathcal{X}_2)$ is equal to*

$$\begin{cases} (\mathbb{L} + \frac{1}{4}(n-1)(m-1)(\mathbb{L}-2))T + \frac{1}{4}(n-1)(m-1)(\mathbb{L}-2))N, & n,m \text{ odd.} \\ (\mathbb{L} + \frac{1}{4}(n-2)(m-1)(\mathbb{L}-2) + \frac{1}{2}(m-1)(\mathbb{L}-1))T + \\ \qquad +(\frac{1}{4}(n-2)(m-1)(\mathbb{L}-2) - \frac{1}{2}(m-1)(\mathbb{L}-1))N, & n \text{ even, } m \text{ odd.} \end{cases}$$

where T is the trivial representation and N is the non-trivial one.

Proof. The character variety $\mathcal{X}_2$ is described in [14] by finding a set of equations satisfied by the traces of the matrices of the images by the representation. In [16] the same variety $\mathcal{X}_2$ is described by a geometric method based on the study of eigenvectors and eigenvalues of the matrices. The variety $\mathcal{X}_2$ consists of the following irreducible components: one component consisting of reducible representations, isomorphic to $\mathbb{C}$; and $(n-1)(m-1)/2$ components forming the irreducible locus, each of them isomorphic to $\mathbb{C} - \{0,1\}$. Therefore the motive is $[\mathcal{X}_2] = \mathbb{L} + \frac{1}{2}(n-1)(m-1)(\mathbb{L}-2)$.

As described in [17], the $\mathrm{PGL}(2,\mathbb{C})$-character variety $\bar{\mathcal{X}}_2$ consists of: one component consisting of reducible representations, isomorphic to $\mathbb{C}$; $[\frac{n-1}{2}][\frac{m-1}{2}]$ components of the irreducible locus, each of them isomorphic to $\mathbb{C} - \{0,1\}$; and if n is even and m is odd, $(m-1)/2$ components of the irreducible locus, each of them isomorphic to $\mathbb{C}^*$ (the case m even and n odd is analogous). Note that we can always assume, by swapping n,m if necessary, that m is odd. Therefore we have that for m,n odd, $[\bar{\mathcal{X}}_2] = \mathbb{L} + \frac{1}{4}(n-1)(m-1)(\mathbb{L}-2)$. For n even and m odd, we have $[\bar{\mathcal{X}}_2] = \mathbb{L} + \frac{1}{4}(n-2)(m-1)(\mathbb{L}-2) + \frac{1}{2}(m-1)(\mathbb{L}-1)$.

By (2.7),

$$h_{\mu_2}(\mathcal{X}_2) = [\bar{\mathcal{X}}_2]T + ([\mathcal{X}_2] - [\bar{\mathcal{X}}_2])N\,,$$

where T is the trivial representation, and N is the non-trivial representation ($T = e_0$ and $N = e_1$ in the notation of Section 2). The result follows. □

Now we move to the description of the $\mathrm{SL}(3,\mathbb{C})$-character variety $\mathcal{X}_3$. The following description appears in [17, Sections 8 and 10].

Proposition 3.2 *The components of $\mathcal{X}_3$ are the following:*

- *The component of totally reducible representations, isomorphic to $\mathbb{C}^2$.*

- $[\frac{n-1}{2}][\frac{m-1}{2}]$ *components of partially reducible representations, each isomorphic to* $(\mathbb{C}-\{0,1\})\times\mathbb{C}^*$.
- *If n is even, there are $(m-1)/2$ extra components of partially reducible representations, each isomorphic to* $\{(u,v)\in\mathbb{C}^2|v\neq 0, v\neq u^2\}$. *(The case m even and n odd is analogous.)*
- $\frac{1}{12}(n-1)(n-2)(m-1)(m-2)$ *componens of irreducible representations, of maximal dimension 4, which are isomorphic to $\mathcal{M}/(T\times_{\mathbb{C}^*}T)$, where $\mathcal{M}\subset \mathrm{GL}(3,\mathbb{C})$ are the stable points for the $(T\times_{\mathbb{C}^*}T)$-action (here T are the diagonal matrices acting by multiplication on* $\mathrm{GL}(3,\mathbb{C})$ *on the left and on the right).*
- $\frac{1}{2}(n-1)(m-1)(n+m-4)$ *components of irreducible representations, each isomorphic to* $(\mathbb{C}^*)^2-\{x+y=1\}$.

From here, we can read off the motive of the character variety $\mathcal{X}_3$ (cf. [17, Theorem 8.3]):

$$\begin{aligned}
[\mathcal{X}_3] &= \frac{1}{12}(n-1)(n-2)(m-1)(m-2)(\mathbb{L}^4+4\mathbb{L}^3-3\mathbb{L}^2-15\mathbb{L}+12)\\
&\quad+\mathbb{L}^2+\frac{1}{4}(n-1)(m-1)(\mathbb{L}^2-3\mathbb{L}+2)\\
&\quad+\frac{1}{2}(n-1)(m-1)(n+m-4)(\mathbb{L}^2-3\mathbb{L}+3),\quad m,n \text{ odd},\\
[\mathcal{X}_3] &= \frac{1}{12}(n-1)(n-2)(m-1)(m-2)(\mathbb{L}^4+4\mathbb{L}^3-3\mathbb{L}^2-15\mathbb{L}+12)\\
&\quad+\mathbb{L}^2+\frac{1}{4}(n-2)(m-1)(\mathbb{L}^2-3\mathbb{L}+2)+\frac{1}{2}(m-1)(\mathbb{L}^2-2\mathbb{L}+1)\\
&\quad+\frac{1}{2}(n-1)(m-1)(n+m-4)(\mathbb{L}^2-3\mathbb{L}+3),\quad n \text{ even}, m \text{ odd}.
\end{aligned}$$

Now we describe the $\mathrm{PGL}(3,\mathbb{C})$-character varieties $\bar{\mathcal{X}}_3$.

Proposition 3.3 *The components of the* $\mathrm{PGL}(3,\mathbb{C})$*-character variety* $\bar{\mathcal{X}}_3$ *are:*

- *The component of totally reducible representations, which is isomorphic to* $\mathbb{C}^2/\mu_3\cong\{(x,y,z)\in\mathbb{C}^3 \mid x\,y=z^3\}$.
- $[\frac{n-1}{2}][\frac{m-1}{2}]$ *components of partially reducible representations, each isomorphic to* $(\mathbb{C}-\{0,1\})\times\mathbb{C}^*$.
- *When n is even, there are $(m-1)/2$ additional components of partially reducible representations, each isomorphic to* $\{(u,v)\in\mathbb{C}^2|v\neq 0, v\neq u^2\}$.
- *When $m,n\notin 3\mathbb{Z}$, there are the following components of irreducible representations:*
  - $(n-1)(m-1)(n+m-4)/6$ *components isomorphic to* $(\mathbb{C}^*)^2-\{x+y=1\}$
  - *and $(m-1)(m-2)(n-1)(n-2)/36$ components of maximal dimension isomorphic to* $\mathcal{M}/(T\times_{\mathbb{C}^*}T)$.

- *When $n \in 3\mathbb{Z}$, there are the following components of irreducible representations:*
  - $(m-1)(mn+n^2-5n-m+2)/6$ *components isomorphic to* $(\mathbb{C}^*)^2-\{x+y=1\}$,
  - $m-1$ *components isomorphic to* $\{(x,y,z)\in\mathbb{C}^3 \mid xy=z^3,\ x+y+3z\neq 1\}$,
  - $(m-1)(m-2)n(n-3)/36$ *components of maximal dimension isomorphic to* $\mathcal{M}/(T\times_{\mathbb{C}^*}T)$,
  - *and* $(m-1)(m-2)/6$ *components of maximal dimension isomorphic to* $\mathcal{M}/(T\times_{\mathbb{C}^*}T\rtimes\mu_3)$, *where* μ_3 *acts by cyclic permutation of columns in* $\mathcal{M}$.

  The case $m\in 3\mathbb{Z}$ is symmetric.

The motive of the character variety $\bar{\mathcal{X}}_3$ is as follows (see [17, Corollary 10.3]):

- If $n,m\equiv 1,5 \pmod 6$, then $[\bar{\mathcal{X}}_3]=P_0+\frac{1}{36}(m-1)(m-2)(n-1)(n-2)P_1+\frac{1}{6}(n-1)(m-1)(n+m-4)P_3+\frac{1}{4}(n-1)(m-1)P_5$.
- If $n\equiv 2,4 \pmod 6$, $m\equiv 1,5 \pmod 6$, then $[\bar{\mathcal{X}}_3]=P_0+\frac{1}{36}(m-1)(m-2)(n-1)(n-2)P_1+\frac{1}{6}(n-1)(m-1)(n+m-4)P_3+\frac{1}{4}(n-2)(m-1)P_5+\frac{1}{2}(m-1)P_6$.
- If $n\equiv 3 \pmod 6$, $m\equiv 1,5 \pmod 6$, then $[\bar{\mathcal{X}}_3]=P_0+\frac{1}{36}(m-1)(m-2)n(n-3)P_1+\frac{1}{6}(m-1)(m-2)P_2+\frac{1}{6}(m-1)(mn+n^2-5n-m-2)P_3+(m-1)P_4+\frac{1}{4}(n-1)(m-1)P_5$.
- If $n\equiv 0 \pmod 6$, $m\equiv 1,5 \pmod 6$, then $[\bar{\mathcal{X}}_3]=P_0+\frac{1}{36}(m-1)(m-2)n(n-3)P_1+\frac{1}{6}(m-1)(m-2)P_2+\frac{1}{6}(m-1)(mn+n^2-5n-m-2)P_3+(m-1)P_4+\frac{1}{4}(n-2)(m-1)P_5+\frac{1}{2}(m-1)P_6$.
- If $n\equiv 2,4 \pmod 6$, $m\equiv 3 \pmod 6$, then $[\bar{\mathcal{X}}_3]=P_0+\frac{1}{36}m(m-3)(n-1)(n-2)P_1+\frac{1}{6}(n-1)(n-2)P_2+\frac{1}{6}(n-1)(mn+m^2-n-5m-2)P_3+(n-1)P_4+\frac{1}{4}(n-2)(m-1)P_5+\frac{1}{2}(m-1)P_6$.

Here $P_0=\mathbb{L}^2$, $P_1=\mathbb{L}^4+4\mathbb{L}^3-3\mathbb{L}^2-15\mathbb{L}+12$, $P_2=\mathbb{L}^4+2\mathbb{L}^3-3\mathbb{L}^2-\mathbb{L}+4$, $P_3=\mathbb{L}^2-3\mathbb{L}+3$, $P_4=\mathbb{L}^2-\mathbb{L}+1$, $P_5=\mathbb{L}^2-3\mathbb{L}+2$, $P_6=\mathbb{L}^2-2\mathbb{L}+1$.

Now, to compute the μ_3-equivariant motive, we use (2.7):

$$h_{\mu_3}(\mathcal{X}_3)=[\bar{\mathcal{X}}_3]T+\frac{1}{2}([\mathcal{X}_3]-[\bar{\mathcal{X}}_3])(R_1+R_2)$$

where T is the trivial representation of μ_3 and R_1,R_2 are the non-trivial representations (T corresponds to e_0 and R_1,R_2 correspond to e_1,e_2). Note that R_1,R_2 are representations defined over $\mathbb{C}$, but R_1+R_2 is a representation defined over the rationals. Theorem 1.1 follows from this.

References

[1] S. del Baño Rollín, V. Navarro-Aznar: *On the motive of a quotient variety.* Collectanea Math. **49** (1998) 203–226.

[2] F. Bittner: *The universal Euler characteristic for varieties of characteristic zero.* Compositio Math. **140** (2004) 1011–1032.

[3] M. Culler, P. Shalen: *Varieties of group representations and splitting of 3-manifolds* Annals of Math. **117** (1983) 109–146.

[4] F. González-Acuña, J.-M. Montesinos-Amilibia: *On the character variety of group representations in* $\mathrm{SL}(2,\mathbb{C})$ *and* $PSL(2,\mathbb{C})$. Math. Z. **214** (1993) 627–652.

[5] T. Hausel, M. Thaddeus: *Mirror symmetry, Langlands duality and Hitchin systems.* Invent. Math. **153** (2003) 197–229.

[6] M. Heusener, J. Porti: *Representations of knot groups into* $\mathrm{SL}(n,\mathbb{C})$ *and twisted Alexander polynomials.* ArXiv:1406.3705.

[7] N. Hitchin: *The self-duality equations on a Riemann surface.* Proc. London Math. Soc. (3) **55** (1987) 59–126.

[8] S. Lang: *Algebra (Revised Third Edition).* GTM **211**, Springer-Verlag 2005.

[9] S. Lawton: *Minimal affine coordinates for* $\mathrm{SL}(3,\mathbb{C})$*-character varieties of free groups.* J. Algebra **320** (2008) 3773–3810.

[10] S. Lawton, V. Muñoz: *E-polynomial of the* $\mathrm{SL}(3,\mathbb{C})$*-character variety of free groups.* ArXiv:1405.0816.

[11] M. Logares, V. Muñoz, P. Newstead: *Hodge polynomials of* $\mathrm{SL}(2,\mathbb{C})$*-character varieties for curves of small genus.* Rev. Mat. Complut. **26** (2013) 635–703.

[12] A. Lubotzky, A. Magid: *Varieties of representations of finitely generated groups.* Memoirs AMS **58**, 1985.

[13] Y. Manin: *Correspondences, motifs and monoidal transformations.* Math. USSR Sb. **6** (1968) 439–470.

[14] J. Martín-Morales, A-M. Oller-Marcén: *On the varieties of representations and characters of a family of one-relator subgroups.* Topol. Appl. **156** (2009) 2376–2389.

[15] J. Martínez, V. Muñoz: *E-polynomials of the* $\mathrm{SL}(2,\mathbb{C})$*-character varieties of surface groups.* Arxiv:1407.6975.

[16] V. Muñoz: *The* $\mathrm{SL}(2,\mathbb{C})$*-character varieties of torus knots*, Rev. Mat. Complut. **22** (2009) 489–497.

[17] V. Muñoz, J. Porti: *Geometry of the* $\mathrm{SL}(3,\mathbb{C})$*-character variety of torus knots.* Arxiv.org:1409.4784.

[18] D. Rolfsen: *Knots and links.* Publish or Perish, 1990.

Spectral limits of semiclassical commuting self-adjoint operators

Álvaro Pelayo* **and San Vũ Ngọc†**

Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112, USA
alpelayo@math.ucsd.edu

Institut de Recherches Mathématiques de Rennes
Université de Rennes 1
Campus de Beaulieu
F-35042 Rennes cedex, France
san.vu-ngoc@univ-rennes1.fr

Dedicated to Professor J. M. Montesinos Amilibia, with admiration.

ABSTRACT

Using an abstract notion of semiclassical quantization for self-adjoint operators, we prove that the joint spectrum of a collection of commuting semiclassical self-adjoint operators converges to the classical spectrum given by the joint image of the principal symbols, in the semiclassical limit. This includes Berezin-Toeplitz quantization and certain cases of $\hbar$-pseudodifferential quantization, for instance when the symbols are uniformly bounded, and extends a result by L. Polterovich and the authors. In the last part of the paper we review the recent solution to the inverse problem for quantum integrable systems with periodic Hamiltonians, and explain how it also follows from the main result in this paper.

2010 Mathematics Subject Classification: 53D05, 37J10, 47A10, 58J53, 81R12.
Key words: spectral, symplectic, inverse problem, microlocal, semiclassical.

---

*Author is partially supported by NSF CAREER grant DMS-1518420, Lebesgue Chair 2015, and Severo Ochoa grant Sev-2011-0087.

†Author is partially supported by the Institut Universitaire de France and the Lebesgue Center (ANR Labex LEBESGUE) .

1. Introduction

In inverse spectral problems one tries to recover geometric (or "classical") information from the spectrum of a "quantum" operator. For instance, does the spectrum of the Laplacian on a bounded euclidean domain completely determine the geometry of the domain ? The problem goes back to S. Bochner and H. Weyl [41, 42] in the late nineteenth and early twentieth century, and was made popular, in the context of Riemannian geometry, in M. Kac's famous article on "can you hear the shape of a drum", [24], who attributes the origin of the question to Bochner.

In this paper we will deal with a quite general setting of semiclassical self-adjoint operators. Roughly speaking, a quantum operator will be a family of operators $(T_\hbar)$ depending on a small real parameter $\hbar > 0$ reminiscent of the Planck constant. To each such operator, one defines its "classical limit" to be a smooth function on a smooth manifold (the phase space), called the *principal symbol* of the operator. The semiclassical inverse problem is then the following.

Question 1.1 (Semiclassical Inverse Spectral Problem) *Given the semiclassical joint spectrum*

$$(X_\hbar)_{\hbar>0} \subset \mathbb{R}^d$$

of a quantum system of commuting semiclassical operators

$$T_1 := (T_{1,\hbar})_{\hbar>0}, \dots, T_d := (T_{d,\hbar})_{\hbar>0},$$

how much can one recover about the classical system given by the principal symbols

$$f_1, \dots, f_d$$

of $T_1, \dots, T_d$*?*

Of course, a complete answer to this question would be to fully recover the principal symbols $f_1, \dots, f_d$ themselves.

One can hope to obtain such general results by combining the use of microlocal and symplectic techniques in the spirit of Duistermaat, Helffer, Hörmander, Sjöstrand, etc. However, to date only a handful of results are known in this direction, see [6, 7, 39, 5, 25, 26]. If one restricts the class of operators to be Schrödinger operators, whose principal symbol is of the form $\xi^2 + V(x)$, then the question amounts to recovering the potential V; this has attracted a lot of mathematicians, and is still an active area of research, see [23, 8, 21].

In this paper we prove a general result giving a partial answer to Question 1.1, inspired by previous works of Colin de Verdiere [6, 7], Polterovich, and the authors [31], which says that even though we do not know how to recover the principal symbols themselves, we can recover the closure of their joint image, which is a subset of the affine space $\mathbb{R}^d$. This gives a rigorous proof of the quantum mechanical principle that says that: "in the high frequency limit $\hbar \to 0$, the spectrum of a quantum system converges to the numerical range of its associated classical system".

Theorem 1.2 *Let $I \subset (0,\, 1]$ be a set with a limit point at 0. Then the limit set of the joint spectrum of a family of pairwise commuting self-adjoint semiclassical operators*

$$T_1 := (T_{1,\hbar})_{\hbar \in I},\, \ldots,\, T_d := (T_{d,\hbar})_{\hbar \in I}$$

is the classical spectrum $\mathcal{S} \subset \mathbb{R}^d$ of $T_1, \ldots, T_d$, that is, the closure of the joint image of the principal symbols of $T_1, \ldots, T_d$.

An illustration of the convergence statement in Theorem 1.2 is depicted in Figure 1, which shows the joint spectrum of the "normalized" Quantum Spherical Pendulum[1]

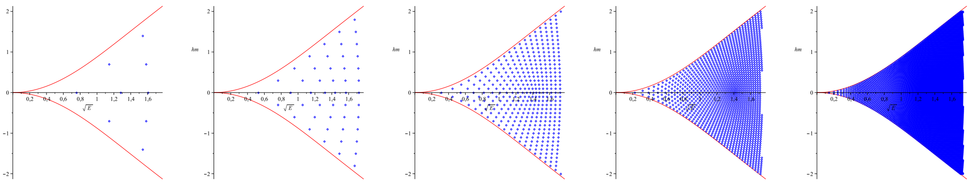


Figure 1: The dots in the figures form the semiclassical joint spectrum of the "normalized" Quantum Spherical Pendulum for the values of the Planck constant: $\hbar = 0.7, 0.5, 0.3, 0.05, 0.02$. As $\hbar \to 0$, the semiclassical joint spectrum fills the inside of the red curve, which is the boundary of the classical spectrum of the system; this gives an illustration of the convergence stated in Theorem 1.2.

In [31] an analogous statement was proved, but taking the convex full on both the quantum and the classical spectrum. We achieve this improvement by introducing a new hypothesis, which takes the form of the following seemingly simple axiom for the abstract semiclassical quantization: for any symbol f, one should have

$$\|\mathrm{Op}_\hbar(f)^2 \; - \mathrm{Op}_\hbar(f^2)\| = \mathcal{O}(\hbar).$$

where $\mathrm{Op}_\hbar$ denotes the quantization operation. We refer to Theorem 4.1 for a detailed version of the above statement, to Definition 2.5 for the abstract notion of semiclassical operators we use, and the upcoming sections for the necessary preliminaries. The abstract notion we use, and hence the theorem, apply to Berezin-Toeplitz operators on compact manifolds, and certain classes of pseudodifferential operators (this is explained in Remark 2.1), for instance those with uniformly bounded derivatives.

As we will explain, Theorem 4.1 implies, in combination with a theorem of Atiyah--Guillemin-Sternberg and Delzant, a solution to the inverse problem for quantum toric integrable systems, which recovers a recent result of Charles and the authors [5]. This result was proved again shortly after by Polterovich and the authors [31] with a different method, which is in fact the one which serves as inspiration for Theorem 4.1, in combination with ideas introduced by Le Floch and the authors in [26].

[1] Instead of the standard energy and momentum operators, the energy is replaced by its square root, in order to obtain symbols which are both (asymptotically) homogeneous of degree one.

2. Semiclassical operators and an abstract semiclassical quantization

We review Berezin-Toeplitz quantization, $\hbar$-pseudodifferential quantization, and then introduce an abstract notion of semiclassical quantization which includes the former, and certain classes of the latter. This abstract notion is inspired by, and extends, a notion introduced by Polterovich and the authors [31] and as we will see in Section 4 it allows us to prove a stronger convergence result in certain cases.

2.1. Berezin-Toeplitz operators

The microlocal analysis of Berezin-Toeplitz operators is rapidly evolving nowadays, see for instance [3, 9, 10, 11, 12, 28, 35] following the pioneer work of Boutet de Movel and Guillemin [4].

Let us recall the basic facts we need on connections of Hermitian line bundles (a good reference for this material are Duistermaat's notes [17]). With the help of these facts we will introduce a fundamental notion in both geometry and analysis, that of a prequantum line bundle.

Let M be a smooth manifold. Let $\mathcal{L} \to M$ be a Hermitian line bundle over M. That is, $\mathcal{L} \to M$ is a complex line bundle over M which is endowed with a Hermitian metric. Denote by $\mathrm{C}^\infty(M, \mathcal{L})$ the space of smooth sections of this bundle, and by $\Omega^1(M, \mathcal{L})$ the space of smooth $\mathcal{L}$-valued 1-forms. A *connection of* $\mathcal{L}$ is a linear operator $\nabla : \mathrm{C}^\infty(M, \mathcal{L}) \to \Omega^1(M, \mathcal{L})$ which satisfies Leibniz's rule, that is,

$$\nabla(fs) = \mathrm{d}f \otimes s + f\nabla s$$

for all smooth functions $f \in \mathrm{C}^\infty(M)$ and all smooth sections $s \in \mathrm{C}^\infty(M, \mathcal{L})$.

Let X be a smooth vector field on M. The *covariant derivative of the section* s of $\mathcal{L}$ with respect to the vector field X is given by the formula $\nabla_X s = \nabla s(X)$. The smooth 2-form R of M defined by the equation

$$R(X, Y) = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]}$$

for any vector fields X, Y of M is called the *curvature of the connection.* Let $(\cdot, \cdot)$ denote the Hermitian scalar product. We say that the connection is *compatible with the Hermitian structure* if $\mathrm{d}(s, t) = (\nabla s, t) + (s, \nabla t)$, for any smooth sections s and t of the line bundle $\mathcal{L}$. In this case, $R = \frac{1}{\mathrm{i}}\omega$ where ω is real-valued. Throughout the present paper all connections considered are implicitly assumed to be compatible with the metric.

Assume that $\frac{1}{\mathrm{i}}\omega$ is the curvature of a Hermitian line bundle connection. Then the cohomology class of the form $\omega/2\pi$ is integral, that it, it lies in the image of the canonical homomorphism

$$H^2(M, \mathbb{Z}) \to H^2(M, \mathbb{R}).$$

Conversely, for any smooth 2-form $\omega \in \Omega^2(M, \mathbb{R})$ for which $[\omega]/2\pi$ is integral there is a Hermitian line bundle $\mathcal{L} \to M$ endowed with a connection ∇ whose curvature is

$\frac{1}{\mathrm{i}}\omega$. Moreover, the line bundle $\mathcal{L}$ and ∇ are unique up to isomorphisms. For a proof of these results, we refer the reader to [17, Theorem 10.1] or [16, Section 15.3]. Now we are ready to recall the following essential definition.

Definition 2.1 *Let (M, ω) be a symplectic manifold.*

- *A* prequantum bundle *on M is a Hermitian line bundle $\mathcal{L} \to M$ with a connection of curvature $\frac{1}{\mathrm{i}}\omega$.*

  In this case we say that the symplectic manifold (M, ω) is prequantizable*.*

- *A* prequantum bundle automorphism *is a vector bundle automorphism of $\mathcal{L} \to M$ which preserves both the metric and the connection.*

Let us know consider a prequantum line bundle $\mathcal{L} \to M$ over the symplectic manifold M. Let G be a Lie group with Lie algebra $\mathfrak{g}$. Assume that G acts on $\mathcal{L}$ by prequantum bundle automorphisms, as defined above. This G-action lifts a G-action on M. For $X \in \mathfrak{g}$ let $X^\sharp$ denote the infinitesimal action of X on M. The latter G-action is Hamiltonian with momentum map F given by the following condition: the action induced by $\mathfrak{g}$ on $\mathrm{C}^\infty(M, \mathcal{L})$ is given by the Kostant-Souriau operators

$$f \mapsto \nabla_{X^\sharp} f + \mathrm{i}\langle F, X\rangle f, \qquad X \in \mathfrak{g}. \tag{2.1}$$

and ∇ denotes the covariant derivative of the prequantum bundle (cf. [16, Proposition 15.2]). If both the Lie group G and the manifold M are connected, then the G-action on $\mathcal{L}$ is conversely determined by the action on M and by the momentum map F. It is important to notice that we cannot obtain every momentum map generating a given action in this manner (such momentum maps correspond to the Lie algebra representations on the prequantum bundle by means of (2.1).

Suppose that (M, ω) is a prequantizable closed (that is both compact and with no boundary) symplectic manifold. Let $\mathcal{L} \to M$ be a prequantum line bundle, and assume that M admits a complex structure that is compatible with the symplectic form ω. In fact, (M, ω) is a Kähler manifold. The holomorphic structure of $\mathcal{L} \to M$ is uniquely determined by the compatibility condition with the connection.

Consider a positive integer $k = 1/\hbar$ and let $\mathcal{L}^k$ denote the kth tensor power of the line bundle $\mathcal{L}$. We write

$$\mathcal{H}_\hbar := H^0(M, \mathcal{L}^k)$$

for the space of holomorphic sections of $\mathcal{L}^k$. The space $\mathcal{H}_\hbar$ is a finite dimensional subspace of the Hilbert space $L^2(M, \mathcal{L}^k)$ (this is because M is a closed manifold). If λ denotes the Liouville measure of M, the scalar product is given by integration of the Hermitian (pointwise) scalar product of sections against λ.

Definition 2.2 *Let $\Pi_\hbar$ be the surjective orthogonal projector*

$$\mathrm{L}^2(M, \mathcal{L}^k) \to \mathcal{H}_\hbar.$$

- *A semiclassical* Berezin-Toeplitz operator *is any sequence of the form*

$$T := (T_\hbar := \Pi_\hbar f(\cdot,\, k)\colon \mathcal{H}_\hbar \to \mathcal{H}_\hbar)_{\hbar=1/k,\ k\in\mathbb{N}^*}$$

where the multiplication operator $f(\cdot,\, k)$ *is a sequence in* $\mathrm{C}^\infty(M)$ *with an asymptotic expansion*

$$f_0 + k^{-1} f_1 + k^{-2} f_2 + \cdots$$

for the C^∞ *topology.*

- *The first coefficient* f_0 *is called the* principal symbol *of* T.

2.2. Pseudodifferential operators

$\hbar$-pseudodifferential operators acting on the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^n)$ give a semiclassical quantization of the manifold $M = \mathbb{R}^{2n}$; this is a semiclassical version of the one given by homogeneous pseudodifferential operators, see for instance [14] or [43].

Let $\mathcal{A}_0$ be the Hörmander class whose elements are the functions f in the space $\mathrm{C}^\infty(\mathbb{R}^{2n}_{(x,\xi)})$ such that the following holds: there is $m \in \mathbb{R}$ for which

$$|\partial^\alpha_{(x,\xi)} f| \leqslant C_\alpha \langle (x,\xi)\rangle^m \tag{2.2}$$

for every $\alpha \in \mathbb{N}^{2n}$ (here the notation $\langle z\rangle$ stands for $(1+|z|^2)^{1/2}$). Symbolic calculus for pseudodifferential operators is known to hold in the case when the symbols of the operators are in $\mathcal{A}_0$ (for instance).

Definition 2.3 *Let* $f \in \mathcal{A}_0$. *The* Weyl quantization of f *is given on the Schwartz space* $\mathcal{S}(\mathbb{R}^n)$ *by the expression:*

$$(\mathrm{Op}_\hbar(f)u)(x) := \frac{1}{(2\pi\hbar)^n}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \mathrm{e}^{\frac{\mathrm{i}}{\hbar}((x-y)\cdot\xi)} f(\tfrac{x+y}{2},\xi)u(y)\,\mathrm{d}y\,\mathrm{d}\xi.$$

This definition is commonly used to define $\hbar$-pseudodifferential operators on $\mathbb{R}^n$. It can also give a semiclassical quantization of a cotangent bundle $M = T^*X$, where X is a smooth n-dimensional closed manifold with a smooth density, as follows. Let X be covered by a collection of smooth charts

$$\Big\{U_1, \ldots, U_N\Big\},$$

where each U_i is over a convex bounded domain of the Euclidean space $\mathbb{R}^n$ (equipped with the Lebesgue measure). By standard manifold theory, there exists a partition of unity

$$\chi_1^2, \ldots, \chi_N^2$$

which is subordinated to the cover $\{U_1, \ldots, U_N\}$. In this case, $\mathcal{A}_0$ is the space of functions $f \in \mathrm{C}^\infty(\mathrm{T}^*X)$ satisfying, for all $(x,\xi) \in \mathrm{T}^*X$, $\alpha \in \mathbb{N}^n$, and for some $m \in \mathbb{R}$, the condition:

$$\left|\partial_x^\alpha \partial_\xi^\beta f(x,\xi)\right| \leqslant C_\alpha \langle \xi \rangle^{m-|\beta|} \tag{2.3}$$

Recall Definition 2.3 and let $\mathrm{Op}_\hbar^j(f)$ be the Weyl quantization in U_j. Define:

$$\mathrm{Op}_\hbar(f)u := \sum_{j=1}^N \chi_j \cdot \mathrm{Op}_\hbar^j(f)(\chi_j u)\,, \quad u \in \mathrm{C}^\infty(X)\,,$$

which is a pseudodifferential operator on X. The principal symbol of this operator is the smooth function

$$f := \sum_{i=1}^N f\chi_j^2.$$

Definition 2.4 *Let X be either $\mathbb{R}^n$, or a closed manifold, as above. Let $(f_\hbar)_{\hbar\in(0,1]}$ be a family of elements of $\mathcal{A}_0$ such that the estimate* (2.2) *(in the case $X = \mathbb{R}^n$) or* (2.3) *(if X is a closed manifold) holds uniformly for $\hbar \in (0,1]$. Then the family*

$$T := (\mathrm{Op}_\hbar(f_\hbar))_{\hbar\in(0,\,1]}$$

is called a semiclassical $\hbar$-pseudodifferential operator on X.

The above definition may also be made for a subset $I \subset (0,\,1]$ which has a limit point at 0, for instance

$$I = \left\{\frac{1}{k} \mid k \in \mathbb{N}^*\right\}.$$

2.3. Abstract semiclassical quantization

The results presented in this paper hold for both pseudodifferential and Berezin-Toeplitz quantization. In fact, they only require a few key properties, and it is interesting to state them in an abstract way, as follows.

Let $I \subset (0,1]$ be a set that accumulates at 0. Suppose that M is a connected manifold (closed or open) and let $\mathcal{A}_0$ be a subalgebra of the algebra of smooth functions $\mathrm{C}^\infty(M;\mathbb{R})$ containing all constants as well as all compactly supported functions.

For a complex Hilbert space $\mathcal{H}$ we denote by $\mathcal{L}(\mathcal{H})$ the set of all linear self-adjoint operators on $\mathcal{H}$ (bounded or unbounded). The following definition is essentially the same as in [31] with the exception of the new Axiom (Q5), which is needed for the proof of our main result.

Definition 2.5 *A* semiclassical quantization *of the pair* $(M, \mathcal{A}_0)$ *is given by:*

- *a family of complex Hilbert spaces $\mathcal{H}_\hbar$, $\hbar \in I$, and*

- *a family of* $\mathbb{R}$*-linear maps* $\mathrm{Op}_\hbar\colon \mathcal{A}_0 \to \mathcal{L}(\mathcal{H}_\hbar)$,

that satisfy the following axioms, where $f, g \in \mathcal{A}_0$*:*

(Q1) $\|\mathrm{Op}_\hbar(1) - \mathrm{Id}\| = \mathcal{O}(\hbar)$ (normalization)*;*

(Q2) *for every function* $f \geqslant 0$ *there is a constant* C_f *for which* $\mathrm{Op}_\hbar(f) \geqslant -C_f\hbar$ (quasi-positivity)*;*

(Q3) *if* $f \in \mathcal{A}_0$ *is that such that* $f \neq 0$ *and also has compact support, then*

$$\liminf_{\hbar \to 0} \|\mathrm{Op}_\hbar(f)\| > 0$$

(non-degeneracy)*;*

(Q4) *if* g *has compact support, then the operator* $\mathrm{Op}_\hbar(f) \circ \mathrm{Op}_\hbar(g)$ *is bounded for every* f*, and*

$$\|\mathrm{Op}_\hbar(f) \circ \mathrm{Op}_\hbar(g) - \mathrm{Op}_\hbar(fg)\| = \mathcal{O}(\hbar),$$

(product formula)*;*

(Q5) *if* $f \in \mathcal{A}_0$*, then*

$$\|\mathrm{Op}_\hbar(f)^2 \ - \mathrm{Op}_\hbar(f^2)\| = \mathcal{O}(\hbar),$$

(square formula).

We say that a manifold is quantizable *if it has a semiclassical quantization.*

Let $\mathcal{A}_I$ be the algebra whose elements are collections $\vec{f} = (f_\hbar)_{\hbar \in I}$, $f_\hbar \in \mathcal{A}_0$, that satisfy that for each $\vec{f}$ there is $f_0 \in \mathcal{A}_0$ such that

$$f_\hbar = f_0 + \hbar f_{1,\hbar}\ , \tag{2.4}$$

where $f_{1,\hbar}$ is uniformly bounded in the parameter $\hbar$ as well as supported in the same compact subset $K(\vec{f})$ of M.

Definition 2.6 *A* semiclassical operator *is an element in the image of the map*

$$Op : \mathcal{A}_I \to \prod_{\hbar \in I} \mathcal{L}(\mathcal{H}_\hbar)$$

defined by

$$\vec{f} = (f_\hbar) \mapsto (\mathrm{Op}_\hbar(f_\hbar))\ .$$

Definition 2.7 *Let* $\vec{f} \in \mathcal{A}_I$*. The function* $f_0 \in \mathcal{A}_0$ *given by* (2.4) *is called the* principal symbol *of* $Op(\vec{f})$.

It follows from the axioms in Definition 2.5 that the principal symbol in Definition 2.7 is uniquely defined, see [31].

Notice that in the definition of semiclassical operators, the manifold M is not required to be symplectic. The following proposition gives the two major examples of semiclassical operators, for which the phase space M is symplectic.

Proposition 2.1 1. *Semiclassical Berezin-Toeplitz operators satisfy the axioms* (Q1–Q5).

2. *Semiclassical pseudodifferential operators which mildly depends on $\hbar$ satisfy the axioms* (Q1–Q4). *Here we say that $f_\hbar \in \mathcal{A}_I$ mildly depends on $\hbar$ if $f_\hbar$ can be written*
$$f_\hbar(x,\xi) = f_0(x,\xi) + \hbar f_{1,\hbar}(x,\xi),$$
where all $f_{1,\hbar}(x,\xi)$ are both uniformly bounded in $\hbar$ as well as compactly supported in the same set.

3. *Semiclassical pseudodifferential operators which are uniformly bounded satisfy the axiom* (Q5). *More precisely, by uniformly bounded we mean that for any $f_\hbar \in \mathcal{A}_I$, and every $\alpha \in \mathbb{N}^{2n}$, there is a constant C_α such that*
$$|\partial^\alpha_{(x,\xi)} f_\hbar(x,\xi)| \leqslant C_\alpha \qquad \forall \hbar \in (0,1].$$

A proof of this can be found in most introductory papers or books on the subject. For instance, for Berezin-Toeplitz operators, one can refer to [3, 9, 10, 11, 12, 28, 35], and for pseudodifferential operators to the books [14] or [43]. Here we do not claim to have optimal hypothesis. For instance, the assumption on mild dependence on $\hbar$ can certainly be weakened.

Remark 2.1 *For pseudodifferential operators, axiom* (Q5) *(square formula) is more restrictive than the others; it does not hold for all classes of symbols. In order to use the results for differential operators like the Laplacian, one would need first a microlocalization estimate in order to truncate the original operators and hence transform them into uniformly bounded pseudodifferential operators. Such a truncation procedure is common in microlocal analysis (see for instance [14, Chapter 10]).*

The following lemma is a consequence of the axioms (Q1–Q4):

Lemma 2.1 ([31, Lemma 11]) *Take any $\vec{f} = (f_\hbar) \in \mathcal{A}_I$ with principal part f_0, and let $(\mathrm{Op}_\hbar(f_\hbar))$ be the corresponding semiclassical operator. Let $\lambda_{\inf}(\hbar) \in [-\infty, +\infty)$ denote the infimum of the spectrum of $\mathrm{Op}_\hbar(f_\hbar)$. Then*
$$\lim_{\hbar\to 0} \lambda_{\inf}(\hbar) = \inf_M f_0. \tag{2.5}$$

3. Joint Spectrum of a family of semiclassical operators

We recall that to any self-adjoint operator A on a Hilbert space, the spectral theorem associates a projector-valued measure μ_A, called the spectral measure, such that

$$A = \int_{\mathbb{R}} td(\mu_A)(t),$$

and whose support is the spectrum of A. A similar theory holds for commuting operators. The self-adjoint operators $S_1, \dots, S_d$ are said to be mutually commuting if their corresponding spectral measures $\mu_1, \dots, \mu_d$ pairwise commute. In this case we may then define the joint spectral measure

$$\mu := \mu_1 \otimes \cdots \otimes \mu_d$$

on $\mathbb{R}^d$. The *joint spectrum* of $(S_1, \dots, S_d)$ is the support of the joint spectral measure, that is:

$$c \in \text{JointSpec}(S_1, \dots, S_d) \iff \forall \epsilon > 0, \quad \mu_1([c_1-\epsilon, c_1+\epsilon]) \circ \cdots \circ \mu_d([c_d-\epsilon, c_d+\epsilon]) \neq 0.$$

In this paper we are interested in the joint spectrum of semiclassical operators, which is defined as follows. For $j \in \{1, \dots, d\}$ let

$$T_j = (T_{j,\hbar})_{\hbar \in I}$$

be semiclassical operators (as in Definitions 2.2 or 2.4) on Hilbert spaces $(\mathcal{H}_\hbar)_{\hbar \in I}$. We assume that for any fixed $h \in I$, the self-adjoint operators $T_{1,\hbar}, \dots, T_{d,\hbar}$ are mutually commuting. For fixed $\hbar$, the *joint spectrum* of $(T_{1,\hbar}, \dots, T_{d,\hbar})$ is as before the support of the joint spectral measure. For instance, if $\mathcal{H}_\hbar$ is finite dimensional (eg. in the case of Berezin-Toeplitz quantization on a closed Kähler manifold), then

$$\text{JointSpec}(T_{1,\hbar}, \dots, T_{d,\hbar})$$

is the set

$$\Big\{(\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d \mid \exists v \neq 0, \; T_{j,\hbar} v = \lambda_j v, \forall j = 1, \dots, d\Big\}.$$

We define the *joint spectrum* of the semiclassical operators $(T_1, \dots, T_d)$ to be the collection of all joint spectra of $(T_{1,\hbar}, \dots, T_{d,\hbar})$, $\hbar \in I$.

4. The inverse problem for commuting operators

4.1. Convergence to classical spectrum

Following the physicists, we use the following definition.

Definition 4.1 *We call* classical spectrum *of $(T_1, \dots, T_d)$ the closure of the image $F(M) \subset \mathbb{R}^d$, where $F = (f_1, \dots, f_d)$ is the map of principal symbols of $T_1, \dots, T_d$.*

In order to state the convergence results for the semiclassical spectrum of a collection of operators, we need to use a notion of limit for subsets of $\mathbb{R}^n$.

Definition 4.2 *Let $(A_\hbar)_{\hbar\in I}$ be a family of subsets of $\mathbb{R}^n$, where $I \subset (0,1]$ is a set which accumulates at 0. The* limit set *of $(A_\hbar)_{\hbar\in I}$ is the subset $A_0 \subset \mathbb{R}^n$ defined by*

$$a \in A_0 \iff \forall \epsilon > 0, \forall \hbar_0 \in I, \exists \hbar \in I, \hbar \leqslant \hbar_0, \text{ such that } A_\hbar \cap B(a,\epsilon) \neq \emptyset.$$

Here $B(a,\epsilon)$ is the euclidean ball around a of radius ϵ.

In the case of uniformly bounded subsets of $\mathbb{R}^n$, the limit set is in fact a limit in the sense of the Hausdorff distance, which we recall now. Let $\|\cdot\|$ be the euclidean norm in $\mathbb{R}^n$. For any $\epsilon > 0$ and any subset X of $\mathbb{R}^n$, we denote by X_ϵ the set

$$\bigcup_{x\in X} \Big\{ m \in \mathbb{R}^n \mid \|x-m\| \leqslant \epsilon \Big\}.$$

The *Hausdorff distance* between two subsets A and B of $\mathbb{R}^n$ is the number

$$\inf \Big\{ \epsilon > 0 \mid \; A \subseteq B_\epsilon \text{ and } B \subseteq A_\epsilon \Big\}.$$

We denote it by $\mathrm{d}_H(A,\, B)$.

Definition 4.3 *Let $(A_\hbar)_{\hbar\in I}$ and $(B_\hbar)_{\hbar\in I}$ be families of uniformly bounded subsets of $\mathbb{R}^n$, where $I \subset (0,1]$ is a set which accumulates at 0.*

- *Fix $N \in \mathbb{N}$. We say that*
$$A_\hbar = B_\hbar + \mathcal{O}(\hbar^N)$$
if there exists a constant $C > 0$ such that $\mathrm{d}_H(A_\hbar,\, B_\hbar) \leqslant C\hbar^N$ for every $\hbar \in I$.
- *We say that*
$$A_\hbar = B_\hbar + \mathcal{O}(\hbar^\infty)$$
if $\mathrm{d}_H(A_\hbar,\, B_\hbar) = \mathcal{O}(\hbar^N)$ for every $N \in \mathbb{N}^*$.
- *Let $A_0 \subset \mathbb{R}^n$. We say that A_0 is a Hausdorff limit of $(A_\hbar)_{\hbar\in I}$ if*
$$\lim_{\hbar\to 0} \mathrm{d}_H(A_\hbar,\, A_0) = 0.$$

Remark 4.1 *Let $(A_\hbar)_{\hbar\in I}$ be a family of uniformly bounded subsets of $\mathbb{R}^n$. The limit set of $(A_\hbar)_{\hbar\in I}$ is always a compact subset $A_0 \subset \mathbb{R}^n$, and then A_0 is a Hausdorff limit of $(A_\hbar)_{\hbar\in I}$. Conversely, if a compact set A_0 is a Hausdorff limit of $(A_\hbar)_{\hbar\in I}$, then it coincides with the limit set of $(A_\hbar)_{\hbar\in I}$.*

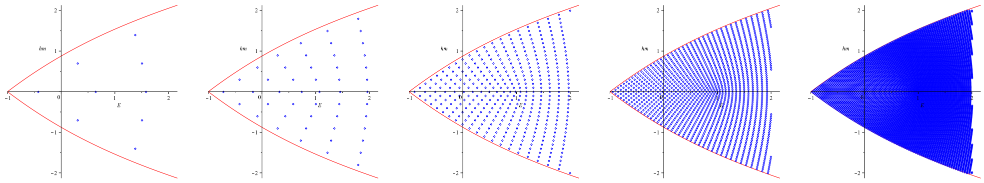


Figure 2: The figure depicts the semiclassical joint spectrum of the Quantum Spherical Pendulum for the following values of the Planck constant: $\hbar = 0.7, 0.5, 0.3, 0.05, 0.02$. As $\hbar \to 0$, the semiclassical joint spectrum fills the inside of the red curve, which is the boundary of the classical spectrum of the system; this gives an illustration of the convergence stated in Theorem 4.1. Notice that, in this figure, the joint spectrum is quite well behaved, because the operators form a completely integrable system. In this case, the joint spectrum is locally diffeomorphic to a lattice, as predicted by the Bohr-Sommerfeld rules, see [40], and in fact much more than the classical spectrum can be recovered from the joint spectrum; see for instance [5, 26].

The following is the main theorem of this paper, which in some cases strengthens previously known theorems (for instance in the case of Berezin-Toeplitz operators, and certain classes of pseudodifferential operators).

Theorem 4.1 *Let $(T_1, \ldots, T_d)$ be a family of pairwise commuting self-adjoint semiclassical operators in the sense of Definition 2.3. Then the limit set of the joint spectrum of $(T_1, \ldots, T_d)$ is the classical spectrum of $(T_1, \ldots, T_d)$.*

Proof. Let μ_j be the spectral measure of T_j, and let $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ be the joint spectral measure of $(T_1, \ldots, T_d)$ on $\mathbb{R}^d$. For any $c = (c_1, \ldots, c_d) \in \mathbb{R}^d$, let $\varphi_c : \mathbb{R}^d \to \mathbb{R}$ be defined by

$$\varphi_c(x_1, \ldots, x_d) := \sum_{i=1}^{d} (x_i - c_i)^2.$$

Then

$$\varphi_c(T_1, \ldots, T_d) = \int_{\mathbb{R}^d} \varphi_c(x) d\mu(x) = \int_{\mathbb{R}} t d((\varphi_c)_* \mu)(t)$$

The last equality implies that the spectrum of $\varphi_c(T_1, \ldots, T_d)$ is the support of $(\varphi_c)_* \mu$.

Now from Axioms (Q1) and (Q5), the principal symbol of $\varphi_c(T_1, \ldots, T_d)$ is

$$\varphi_c(f_1, \ldots, f_d) = \|F - c\|^2$$

where f_i is the principal symbol of T_i. We see that $c \in \overline{F(M)}$ if and only if $\inf \varphi_c(f_1, \ldots, f_d) = 0$. By Lemma 2.1,

$$\inf_M \varphi_c(f_1, \ldots, f_d) = \lim_{\hbar \to 0} \inf \operatorname{Spec}(\varphi_c(T_1, \ldots, T_d)). \tag{4.1}$$

We have $\mathrm{Spec}(\varphi_c(T_1,\dots,T_d)) = \mathrm{supp}((\varphi_c)_*\mu)$. Since φ_c is continuous,

$$\mathrm{supp}((\varphi_c)_*\mu) = \overline{\varphi_c(\mathrm{supp}(\mu))}. \tag{4.2}$$

Assume that c is not in the limit set of the joint spectrum of $(T_1,\dots,T_d)$. Thus there is a small ball around c which is disjoint from $\mathrm{JointSpec}(T_1,\dots,T_d)$ for $\hbar$ small enough, which implies that there is some constant $\epsilon > 0$ such that

$$\inf(\varphi_c(\mathrm{JointSpec}(T_1,\dots,T_d))) > \epsilon.$$

Since $\mathrm{JointSpec}(T_1,\dots,T_d) = \mathrm{supp}(\mu)$, we get in view of (4.2) that

$$\inf(\mathrm{supp}((\varphi_c)_*\mu)) \geqslant \epsilon.$$

Therefore, by Equation (4.1), we get that

$$\inf_M \varphi_c(f_1,\dots,f_d) \geqslant \epsilon.$$

Hence $c \notin \overline{F(M)}$, which says that $\overline{F(M)}$ is contained in the limit set of the joint spectrum.

In fact, all converse implications hold true, which proves the reverse inclusion and hence the theorem. □

Theorem 4.1 shows that the classical spectrum can be recovered from the quantum joint spectrum. In the case of Berezin-Toeplitz operators, this generalizes a result of [31], where the convexity of the classical spectrum was required. For classes of pseudodifferential operators for which Axiom (Q5) does not hold, we cannot apply Theorem 4.1; however, the convex case holds, as we recall below.

Remark 4.2 *We want to emphasize that axiom* (Q5)*, while seemingly simple and quite close indeed to axiom* (Q4)*, gives in fact a great advantage in the form of a rudimentary (polynomial) functional calculus. We conjecture that the strong conclusion of Theorem 4.1, compared to Theorem 4.2 below, could not be obtained by axioms* (Q1–Q4) *alone.*

In what follows we work with $\hbar$-pseudodifferential operators which are not necessarily bounded (see Section 2.2).

Theorem 4.2 ([31]) *Let X be either $\mathbb{R}^n$, or a closed manifold. Let $(T_1,\dots T_d)$ be a family of pairwise commuting self-adjoint semiclassical $\hbar$-pseudodifferential operators on X whose symbols mildly depend*[2] *on $\hbar$. Let $\mathcal{S} \subset \mathbb{R}^d$ be the classical spectrum of $(T_1,\dots,T_d)$. Suppose that $\mathcal{S}$ is a convex set. Then:*

[2]See Proposition 2.1

- *from* JointSpec$(T_1, \ldots, T_d)$ *one can recover* $\mathcal{S}$*;*
- *if moreover each* T_i*,* $1 \leqslant i \leqslant d$*, is bounded, then* $\overline{\mathcal{S}}$ *is the Hausdorff limit, as* $\hbar \to 0$*, of* Convex Hull(JointSpec$(T_{1,\hbar}, \ldots, T_{d,\hbar})$).

For further discussion on these results see [31, 30]. Notice that all the results presented in this paper strongly rely on the self-adjointness of the operators, which ensures a stable behaviour of the spectrum as $\hbar \to 0$. For general non-selfadoint operators, for which there is considerable recent interest (see [38, 36, 37]), similar results can probably be obtained for the semiclassical pseudo-spectrum instead of the spectrum, but to the authors knowledge, this has never been studied in the case of commuting operators. On the other hand, for non-selfadjoint operators that are normal, the stability of the spectrum is expected to hold, see for instance [27].

4.2. The completely integrable case

Let (M, ω) be a $2n$-dimensional symplectic manifold. Given a smooth function $f\colon M \to \mathbb{R}$, we define the *Hamiltonian vector field* $\mathcal{X}_f$ induced by f on M by

$$\omega\,(\mathcal{X}_f,\ \cdot) = -\,\mathrm{d}f.$$

This differential equation (or rather, system of differential equations) is known as Hamilton's equation.

Definition 4.4 *A* classical integrable system *on* (M, ω) *is given by a smooth* $\mathbb{R}^n$*-valued map*

$$F := (f_1,\, \ldots,\, f_n)\colon M \to \mathbb{R}^n$$

such that each component function f_i *is constant along the flow of the vector field* $\mathcal{X}_{f_j}$ *generated by the component* f_j*, for all* i, j*, and the vector fields* $\mathcal{X}_{f_1}, \ldots, \mathcal{X}_{f_n}$ *are linearly independent almost everywhere.*

The first of the conditions in Definition 4.4 can be rephrased as

$$\{f_i, f_j\} = 0$$

for all i, j, where

$$\{f_i, f_j\} := \omega(\mathcal{X}_{f_i}, \mathcal{X}_{f_j})$$

are the so called *Poisson brackets* of f_i and f_j; in this case we say that f_i and f_j are in involution. The integer n (half the dimension of M) in this definition is the largest integer for which the conditions of the definition hold: that is, there is no set of functions

$$f_1, \ldots, f_k,$$

with $k > n$ which satisfies the conditions above and it is in this sense that the word "integrable system" is used.

Remark 4.3 *The symplectic theory of finite dimensional integrable Hamiltonian systems relies on several fundamental results. Liouville-Mineur-Arnold's action-angle theorem [29, 1] is one of the fundamental pieces of the modern theory of integrable systems. Duistermaat [15] described the obstruction to the existence of global-action coordinates in 1980, and this was the starting point of the global symplectic theory of integrable systems. Eliasson [19, 18] proved in the 1980s a major theorem on the linearization of smooth non-degenerate singularities of integrable systems, which continues to be one of the foundational and most useful results of the subjects; the majority of (but not all) results known to date about the general structure of integrable systems, assume that the singularities are non-degenerate.*

In the 1980s the global classification of toric integrable systems of Atiyah, Guillemin--Sternberg, and Delzant opened up the doors and served as inspiration to many authors working on global symplectic invariants of integrable systems. Our next goal is to define toric integrable systems, and the natural transformations between them, and state the classification in the work of Atiyah, Guillemin-Sternberg and Delzant. Let (M, ω) be a $2n$-dimensional symplectic manifold. A smooth map

$$F := (f_1, \ldots, f_n)\colon M \to \mathbb{R}^n$$

on (M, ω) is a *momentum map* for a Hamiltonian n-torus action if each of the Hamiltonian flows

$$t_j \mapsto \varphi_{f_j}^{t_j}$$

of the vector fields $\mathcal{X}_{f_j}$ is periodic of period 1, and all of them pairwise commute, that is,

$$\varphi_{f_j}^{t_j} \circ \varphi_{f_i}^{t_i} = \varphi_{f_i}^{t_i} \circ \varphi_{f_j}^{t_j}$$

so that they define an action of the torus $\mathbb{R}^n/\mathbb{Z}^n$.

Definition 4.5 *We say that a* momentum map *for a Hamiltonian n-torus action is a* toric integrable system*, or simply a* toric system*, if in addition the following conditions hold:*

- *the manifold M is closed and connected;*
- *the action of the torus $\mathbb{R}^n/\mathbb{Z}^n$ is effective.*

The natural transformations between toric integrable systems preserve the toric and the symplectic structure simultaneously, they are precisely given by the following.

Definition 4.6 *Two toric systems (M, ω, F) and (M', ω', F') are* isomorphic *if there exists a symplectomorphism $\varphi : (M, \omega) \to (M', \omega')$ such that $\varphi^* F' = F$.*

Atiyah and Guillemin-Sternberg proved the following influential result (in fact their result applied to much more general momentum maps given by an m-tuple on a $2n$-manifold, where n is not necessarily equal to m and the induced toral action is not necessarily effective):

Theorem 4.3 ([2, 22]) *The image $F(M)$ of a toric system $F\colon M \to \mathbb{R}^n$ is a convex polytope in $\mathbb{R}^n$.*

The set of fixed point (also called elliptic points) of the induced $\mathbb{R}^n/\mathbb{Z}^n$-action is a collection of symplectic submanifolds of M, and its image under F gives a finite collection of points

$$p_1, \ldots, p_k \in \mathbb{R}^n \quad k \geqslant 1.$$

The convex polytope in Theorem 4.3 is precisely the set

$$\Delta = \text{Convex Hull}(\{p_1, \ldots, p_k\}).$$

Shortly after Atiyah and Guillemin-Sternberg proved their theorem, Delzant proved a converse type result, hence giving a classification of toric systems.

Theorem 4.4 ([13]) *The image $F(M)$ of a toric system $F\colon M \to \mathbb{R}^n$ is a Delzant polytope (i.e. rational, simple, and smooth). Moreover, (M, ω, F) is classified, up to isomorphisms, by $F(M)$.*

Theorem 4.4 was generalized in [32, 33] to a class of systems f_1, f_2 on four dimensional manifolds, called *semitoric systems*, in which only f_1 is required to generate a periodic flow.

Definition 4.7 *A* quantum integrable system *is given by a collection of n commuting semiclassical self-adjoint operators*

$$T_1 := (T_{1,\hbar})_{\hbar \in I},\ \ldots,\ T_n := (T_{n,\hbar})_{\hbar \in I}$$

whose principal symbols form a classical integrable system on M.

Definition 4.8 *A quantum integrable system $T_1, \ldots, T_n$ on (M, ω) is* toric *if the principal symbols of $T_1, \ldots, T_n$ are a toric system.*

Remark 4.4 *It is known that not every symplectic manifold has a complex structure or a prequantum line bundle. In the case of toric integrable systems, the situation is better. A toric integrable system does admit a compatible complex structure, which is however not unique. Suppose that $F : M \to \mathbb{R}^n$ is the momentum map and let*

$$\Delta := F(M)$$

be its image in $\mathbb{R}^n$, which is a convex polytope (by Theorem 4.3), say with vertices

$$p_1, \ldots, p_k.$$

Then the system is prequantizable if and only if there is a constant $\ell \in \mathbb{R}^n$ such that

$$p_1 + \ell, \ldots, p_k + \ell \in 2\pi\mathbb{Z}^n.$$

If this holds, the prequantum line bundle is in fact unique, up to isomorphisms.

In the case of Berezin-Toeplitz quantization, it is remarkable that the joint spectrum of a toric system can be completely described, as follows.

Theorem 4.5 ([5]) *Let $T_1, \dots, T_n$ be a Berezin-Toeplitz quantum toric system on a closed manifold M. Then*

$$\mathrm{JointSpec}(T_1, \dots, T_n) = g\Big(\Delta \cap \Big(v + \frac{2\pi}{k}\mathbb{Z}^n\Big);\, k\Big) + \mathcal{O}(k^{-\infty})$$

where:

- *the set $\Delta \subset \mathbb{R}^n$ is the convex polytope $F(M)$ in the Atiyah-Guillemin-Sternberg theorem (Theorem 4.3);*
- *the point $v \in \mathbb{R}^n$ is any vertex of Δ;*
- *$g(\cdot; k) : \mathbb{R}^n \to \mathbb{R}^n$ admits a C^∞-asymptotic expansion*

  $$g(\cdot;\, k) = \mathrm{Id} + k^{-1} g_1 + k^{-2} g_2 + \cdots$$

  where each $g_j : \mathbb{R}^n \to \mathbb{R}^n$ is a smooth function.

Moreover, if the spectral parameter k is large enough then the multiplicity of the eigenvalues of

$$\mathrm{JointSpec}(T_1, \dots, T_n)$$

is precisely equal to 1, and and there is $\delta > 0$ such that if ν is an eigenvalue then the ball $\mathrm{B}(\nu, \delta/k)$ centered at ν or radius δ/k contains precisely only the eigenvalue ν.

As an immediate consequence of this theorem, we have the following.

Corollary 4.1 *Let $T_1, \dots, T_n$ be a quantum toric system on a closed manifold M. Then the joint spectrum of $T_1, \dots, T_n$ modulo $\mathcal{O}(1/k)$ determines the classical integrable system given by principal symbols, up to isomorphisms.*

A quicker alternative proof of Corollary 4.1 was given in [31], and it also follows (in the same way as therein) from Theorem 4.1. Indeed, Theorem 4.1 implies that from the joint spectrum one can recover $\overline{F(M)}$. But we know that $\overline{F(M)} = \Delta$ is the Delzant polytope of the system. Since, by Delzant's theorem 4.4, the polytope is enough to reconstruct the manifold and the moment map F, it follows that the joint spectrum completely determines the classical system. □

To conclude, let us mention that an interesting consequence of the proofs of these results is that any classical toric system can be quantized.

Corollary 4.2 ([5]) *There exists a quantization of any classical toric integrable system. That is, given a classical toric system there is a quantum toric system whose principal symbols are precisely those given by the classical toric system.*

We do not know whether a similar statement holds for more general classes of completely integrable systems. In the analytic case, an algebraic obstruction was constructed in [20].

Acknowledgements. We thank Y. Le Floch and J. Palmer for comments on a preliminary version, and L. Polterovich for useful discussions. Part of this paper was written while the second author was holding the Lebesgue Chair at the Lebesgue Center, during the Thematic Semester in Analysis and PDEs.

References

[1] V. I. Arnold: *A theorem of Liouville concerning integrable problems of dynamics.* Siberian Math. J **4** (1963) 471–474.

[2] M. Atiyah: *Convexity and commuting Hamiltonians.* Bull. London Math. Soc. **14** (1982) 1–15.

[3] D. Borthwick, T. Paul, A. Uribe: *Semiclassical spectral estimates for Toeplitz operators.* Ann. Inst. Fourier (Grenoble) **48** (1998) 1189–1229.

[4] L. Boutet de Monvel, V. Guillemin: *The Spectral Theory of Toeplitz Operators.* Ann. of Math. Studies 99. Princeton University Press, Princeton, NJ, University of Tokyo Press, Tokyo 1981.

[5] L. Charles, Á. Pelayo, S. Vũ Ngọc: *Isospectrality for quantum toric integrable systems* (Dedicated to Peter Sarnak on his 60th Birthday). Annales Sci. Ec. Norm. Sup. **43** (2013) 815–849.

[6] Y. Colin de Verdière: *Spectre conjoint d'opérateurs pseudo-différentiels qui commutent. II. Le cas intégrable.* Math. Z. **171** (1980) 51–73.

[7] Y. Colin de Verdière: *Spectre conjoint d'opérateurs pseudo-différentiels qui commutent. I. Le cas non intégrable.* Duke Math. J. **46** (1979) 169–182.

[8] Y. Colin de Verdière, V. Guillemin: *A semi-classical inverse problem I: Taylor expansions.* In *Geometric aspects of analysis and mechanics,*81–95, Progr. Math. 292, Birkhuser/Springer, New York 2011.

[9] L. Charles: *Berezin-Toeplitz operators, a semi-classical approach.* Comm. Math. Phys. **239** (2003) 1–28.

[10] L. Charles: *Symbolic calculus for Toeplitz operators with half-forms.* Journal of Symplectic Geometry **4** (2006) 171–198.

[11] L. Charles: *Toeplitz operators and Hamiltonian Torus Actions.* J. Functional Analysis **236** (2006) 299–350.

[12] L. Charles: *Semi-classical properties of geometric quantization with metaplectic correction*Comm. Math. Phys. **270** (2007) 445–480.

[13] T. Delzant: *Hamiltoniens périodiques et image convexe de l'application moment. Bull. Soc. Math. France* **116** (1988) 315–339.

[14] M. Dimassi, J. Sjöstrand: *Spectral asymptotics in the semi-classical limit.* London Mathematical Society Lecture Note Series, 268. Cambridge University Press, Cambridge 1999.

[15] J. J. Duistermaat: *On global action-angle variables.* Comm. Pure Appl. Math. **33** (1980) 687–706.

[16] J. J. Duistermaat: *The heat kernel Lefschetz fixed point formula for the spin-c Dirac operators.* Reprint of the 1996 edition. Modern Birkhäuser Classics. Birkhäuser/Springer, New York 2011.

[17] J. J. Duistermaat: *Principal Fiber Bundles.* Notes for *Spring School, June 17-22, 2004, Utrecht.*

[18] L. H. Eliasson: *Normal forms for hamiltonian systems with Poisson commuting integrals – elliptic case.* Comment. Math. Helv. **65** (1990) 4–35.

[19] L. H. Eliasson: *Hamiltonian systems with Poisson commuting integrals.* Ph. D. thesis, University of Stockholm, 1984.

[20] M. Garay, D. van Straten: *Classical and quantum integrability.* Mosc. Math. J. **10** (2010) 519–545.

[21] V. Guillemin, T. Paul: *Some Remarks about Semiclassical Trace Invariants and Quantum Normal Forms.* Comm. Math. Phys. **294** (2009) 1–19.

[22] V. Guillemin, S. Sternberg: *Convexity properties of the moment mapping.* Invent. Math. **67** (1982) 491–513.

[23] H. Hezari: *Inverse spectral problems for Schrödinger operators.* Comm. Math. Phys. (2009)1061–1088.

[24] M. Kac: *Can one hear the shape of a drum?*(Polish) Wiadom. Mat. (2) **13** (1971) 11–35; translated from the English: Amer. Math. Monthly **73** (1966), no. 4, part II, 1–23.

[25] Y. Le Floch: *Inverse spectral theory for 1D Toeplitz operators.* Ph. D. Thesis, University Rennes 1, 2014.

[26] Y. Le Floch, Á. Pelayo, S. Vũ Ngọc. *Semiclassical inverse spectral theory for Jaynes-Cummings type systems.* Math. Annalen, in press.

[27] Y. Le Floch, Á. Pelayo. *Spectral asymptotics of semiclassical unitary operators.* arXiv:1506:02873.

[28] X. Ma; G. Marinescu. *Toeplitz operators on symplectic manifolds.* J. Geom. Anal. **18** (2008) 565–611.

[29] H. Mineur: *Sur les systèmes mécaniques dans lesquels figurent des paramètres fonctions du temps. Étude des systèmes admettant n intégrales premières uniformes en involution. Extension à ces systèmes des conditions de quantification de Bohr-Sommerfeld.* J. École Polytechn. III (1937) (Cahier 1, Fasc. 2 et 3) 173–191, 237–270.

[30] Á. Pelayo: *Symplectic spectral geometry of semiclassical operators.* Bull. Belg. Math. Soc. Simon Stevin **20** (2013) 405–415.

[31] Á. Pelayo, L. Polterovich, S. Vũ Ngọc: *Semiclassical quantization and spectral limits of ℏ-pseudodifferential and Berezin-Toeplitz operators.* Proc. Lond. Math. Soc. **109** (2014) 676–696.

[32] Á. Pelayo, S. Vũ Ngọc: *Semitoric integrable systems on symplectic* 4*-manifolds.* Invent. Math. **177** (2009) 571–597.

[33] Á. Pelayo, S. Vũ Ngọc: *Constructing integrable systems of semitoric type.* Acta Math. **206** (2011) 93–125.

[34] Á. Pelayo, S. Vũ Ngọc: *Semiclassical inverse spectral theory for singularities of focus-focus type.* Comm. Math. Phys. **329** (2014) 809–820.

[35] M. Schlichenmaier: *Berezin-Toeplitz quantization for compact Kähler manifolds. A review of results.* Adv. Math. Phys., Art. ID 927280, 38, 2010.

[36] J. Sjoštrand: *Some results on nonselfadjoint operators: a survey.* In *Further progress in analysis,* 45–74, World Sci. Publ., Hackensack, NJ, 2009.

[37] J. Sjoštrand: *Spectral properties of non-self-adjoint operators.* arxiv:1001.4844v1, 2010.

[38] L. N. Trefethen, M. Embree: *Spectra and pseudospectra. The behavior of nonnormal matrices and operators.* Princeton University Press, 2005.

[39] S. Vũ Ngọc: *Symplectic inverse spectral theory for pseudodifferential operators.* In *Geometric aspects of analysis and mechanics,*353–372, Progr. Math., 292, 2011.

[40] S. Vũ Ngọc: *Bohr-Sommerfeld conditions for integrable systems with critical manifolds of focus-focus type.* Comm. Pure Appl. Math. 53 (2) (2000) 143–217.

[41] H. Weyl: *Über die asymptotische Verteilungder Eigenwerte.* Gott. Nach. (1911) 110–117.

[42] H. Weyl: *Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen.* Math. Ann. **71** (1912) 441–479.

[43] M. Zworski: *Semiclassical analysis.* GSM 138. AMS, Providence, RI, 2012.

Nontrivial twisted Alexander polynomials

Joan Porti[*]

Department de Matemàtiques
Universitat Autònoma de Barcelona
08193 cerdanyola del Vallès, Spain
porti@mat.uab.cat

In memory of Carmen Safont Edo. She taught me geometric topology, as she had learned from her advisor, José Maria Montesinos Amilibia. This article is also devoted to him and to his beautiful mathematical ideas.

ABSTRACT

In this note I study the twisted Alexander polynomial of a hyperbolic knot, twisted by a lift of the holonomy representation in $\mathrm{SL}_2(\mathbb{C})$. Dunfield, Friedl, and Jackson conjectured that its degree is $4g-2$, where g is the genus of the knot. Here I prove that its degree is at least 2, hence it is nontrivial.

2010 Mathematics Subject Classification: 57M25, 57M27.
Key words: Twisted Alexander polynomial, Reidemeister torsion, Lagrangian subspace.

1. Introduction

Let $K \subset S^3$ be a knot and let $\rho : \pi_1(S^3 \setminus K) \to \mathrm{SL}_n(\mathbb{F})$ be a representation of the fundamental group of its complement, where $\mathbb{F}$ denotes a field. Wada [22] and Lin [12] defined the *twisted Alexander polynomial* of K and ρ:

$$\Delta_\rho(t) \in \mathbb{C}[t^{\pm 1}],$$

which is well defined up to multiplication by a factor $\pm t^n$. Conjugate representations yield the same twisted polynomial. When ρ is trivial, one gets the usual (untwisted)

[*]Author partially supported by project MTM2012-34834

Alexander polynomial. Its degree can be defined after taking care of the indeterminacy of the multiplying factor $\pm t^n$:

$$\deg(\Delta_\rho) = \sup\{|m-n| \mid \Delta_\rho(t) = \sum_i a_i t^i,\ a_m, a_n \neq 0\}.$$

A *Seifert surface* is an orientable, compact, and embedded surface $S \subset S^3$ such that $\partial S = K$. The *genus* of the knot is the minimal genus of a Seifert surface:

$$g = g(K) = \min\{\text{genus}(S) \mid S \text{ is a Seifert surface for } K\}.$$

The following upper bound is well known:

$$\deg(\Delta_\rho) \leq n(2g-1),$$

(where n is the dimension of the representation $\rho : \pi_1(S^3 \setminus K) \to \mathrm{SL}_n(\mathbb{F})$) with equality when the knot is fibered, see for instance [5, 13, 7]. In general the equality is not reached.

This paper focuses on hyperbolic knots. The holonomy representation of its (complete) hyperbolic structure

$$\pi_1(S^3 \setminus K) \to \mathrm{Isom}^+(\mathbb{H}^3) \cong \mathrm{PSL}_2(\mathbb{C}) = \mathrm{SL}_2(\mathbb{C})/\{\pm \mathrm{Id}\}$$

lifts to a representation [9, 4, 19]

$$\rho : \pi_1(S^3 \setminus K) \to \mathrm{SL}_2(\mathbb{C}).$$

A choice of such lift is equivalent to the choice of a spin structure [4, 14], so there are two possible choices: if ρ is a lift, then the other one is $(-1)^{\mathrm{ab}}\rho$, where

$$\mathrm{ab} : \pi_1(S^3 \setminus K) \twoheadrightarrow \mathbb{Z}$$

denotes the abelianization map. The corresponding twisted Alexander polynomials differ by a change of sign of the variable:

$$\Delta_{(-1)^{\mathrm{ab}}\rho}(t) = \Delta_\rho(-t).$$

In particular, they have the same degree.

In [6], Dunfield, Friedl, and Jackson computed this polynomial for all hyperbolic knots up to 15 crossings, and stated the following conjecture.

Conjecture 1.1 *[6] Let $K \subset S^3$ be a hyperbolic knot and $\rho : \pi_1(S^3 \setminus K) \to \mathrm{SL}_2(\mathbb{C})$ a lift of the hyperbolic holonomy. Then*

$$\deg(\Delta_\rho) = 2(2g-1).$$

Agol and Dunfield recently proved it for a large class of knots [1], but I am not aware of anybody that has proved that it is nontrivial in general. This is the purpose of this note:

Theorem 1.2 *Let $K \subset S^3$ be a hyperbolic knot and $\rho : \pi_1(S^3 \setminus K) \to \mathrm{SL}_2(\mathbb{C})$ a lift of the hyperbolic holonomy. Then*

$$\deg(\Delta_\rho) \geq 2.$$

In particular Δ_ρ is nontrivial.

The proof applies to other representations $\rho : \pi_1(S^3 \setminus K) \to \mathrm{SL}_2(\mathbb{C})$, provided that ρ satisfies three properties: 1) ρ is acyclic, see Section 2, 2) the restriction of ρ to the Seifert surface is irreducible, and 3) the trace of the longitude is not 2. All these properties of course hold for a lift of the holonomy (the trace of the longitude is -2 [3, 15]).

In this note I provide a proof of Theorem 1.2 using linear algebra. More precisely, I show first that the degree of the twisted Alexander polynomial can be computed from the rank of a map between cohomology with twisted (complex) coefficients. This twisted cohomology is equipped with a symmetric bilinear form, and the image of the relevant map is Lagrangian. Then the argument is based on elementary properties of the geometry of Lagrangian subspaces of $\mathbb{C}^{2n}$. Unfortunately, this argument will never yield a proof of the conjecture, as it does not use the fact that the genus is minimal.

The paper is organized as follows. Section 2 recalls the definitions of twisted cohomology and twisted Alexander polynomial, and the fact that it can be viewed as a Reidemeister torsion. Section 3 is devoted to the relation between the degree and the rank of some maps in twisted cohomology, related to the Seifert surface. The basic properties of the bilinear form on twisted cohomolgy are explained in Section 4 and the discussion on Lagrangian spaces in Section 5. Finally the proof of Theorem 1.2 is given in Section 6.

2. Twisted polynomials

Let $M = S^3 \setminus K$ denote the *knot complement*. Given a representation

$$\rho : \pi_1(M) \to \mathrm{SL}_2(\mathbb{C}),$$

consider the cohomology of M with coefficients twisted by ρ. It is defined using either de Rham or simplicial cohomology. In the de Rham setting, one constructs the bundle

$$E(\rho) = \widetilde{M} \times \mathbb{C}^2 / \pi_1(M),$$

where $\pi_1(M)$ acts on the universal covering $\widetilde{M}$ by deck transformations and on $\mathbb{C}^2$ via ρ. Then the twisted cohomology is the cohomolgy of differential forms valued on this bundle, namely the cohomology of the complex $\Omega(M, E(\rho)) = \Gamma(\Lambda^*(T^*M \otimes E(\rho))$ with the usual differential. In the simplicial setting, one considers a simplicial complex X with underlying manifold the *exterior* of the knot $S^3 \setminus \mathcal{N}(K)$. Then the chain complex considered is

$$C_*(X;\rho) = \hom_{\pi_1(M)}(C_*(\widetilde{X}, \mathbb{Z}), \mathbb{C}^2)$$

where $\pi_1(M)$ acts on the simplicial chains on the universal covering by deck transformations and on $\mathbb{C}^2$ via ρ. Here $\hom_{\pi_1(M)}$ means $\pi_1(M)$-equivariant linear homomorphisms. This cohomology will be denoted by

$$H^*(M;\rho).$$

Similarly, one defines the homology $H_*(M;\rho)$.

It is known by [15] that if ρ is a lift of the holonomy, then it is *acyclic*, namely that

$$H^*(M;\rho) = H_*(M;\rho) = 0.$$

Acyclicity is not essential for the definition of twisted Alexander polynomial, but it simplifies the definition and will be used in the proof of the theorem. So in what follows I restrict to ρ acyclic.

To define twisted Alexander polynomials, one considers the infinite cyclic covering of M

$$\mathbb{Z} \to M_\infty \to M,$$

so that there is an action of $\mathbb{Z}$ on $H^*(M_\infty;\rho)$ and on $H_*(M_\infty;\rho)$, again by deck transformations. By letting t denote the generator of $\mathbb{Z}$, the (co-)homology groups $H^*(M_\infty;\rho)$ and $H_*(M_\infty;\rho)$ can be viewed as $\mathbb{C}[t^{\pm 1}]$-modules, where $\mathbb{C}[t^{\pm 1}]$ denotes the ring of Laurent polynomials. Since ρ is acyclic, it can be proved that $H_i(M_\infty;\rho) = H^i(M_\infty;\rho) = 0$, except for $i = 1$ [20, 17]. In addition, $H^1(M;\rho)$ is a $\mathbb{C}[t^{\pm 1}]$-torsion module, and the twisted Alexander polynomial is defined as the order of this module [20]. In other words, since $\mathbb{C}[t^{\pm 1}]$ is a PID, one may write

$$H^1(M_\infty;\rho) \cong \mathbb{C}[t^{\pm 1}]/\Delta_1(t) \oplus \cdots \oplus \mathbb{C}[t^{\pm 1}]/\Delta_k(t).$$

Definition *For M and ρ as above, the* twisted Alexander polynomial *is the product*

$$\Delta_\rho(t) = \Delta_1(t) \cdots \Delta_k(t),$$

which is well defined up to a factor $c\,t^k$, with $c \in \mathbb{C}^*$ and $k \in \mathbb{Z}$.

Notice that the indeterminacy factors are precisely the units in $\mathbb{C}[t^{\pm 1}]$. The definition given by Wada has only $\pm t^k$ as indeterminacy.

In [18] Milnor proved that the untwisted Alexander polynomial is a Reidemeister torsion. This was generalized by Kitano in [11] for twisted Alexander polynomials. I am not going to define Reidemeister torsion here, just use its properties. The main references for that are [16, 20, 21]. To view the twisted Alexander polynomial as a Reidemeister torsion, I need to consider the abelianization map

$$\mathrm{ab} : \pi_1(M) \to \pi_1(M)/[\pi_1(M), \pi_1(M)] \cong \mathbb{Z},$$

the corresponding polynomial representation

$$\begin{array}{rcl} \varphi : \pi_1(M) & \to & \mathbb{C}[t^{\pm 1}] \\ \gamma & \mapsto & t^{\mathrm{ab}(\gamma)} \end{array}$$

and the tensor product:

$$\rho \otimes \varphi : \pi_1(M) \to \mathrm{SL}_2(\mathbb{C}) \otimes \mathbb{C}[t^{\pm 1}] \subset \mathrm{GL}_2(\mathbb{C}[t^{\pm 1}]).$$

The Reidemeister torsion is denoted by

$$\tau(M; \rho \otimes \varphi) \in \mathbb{C}[t^{\pm 1}]$$

and it is defined up to multiplication by a factor t^k (there is no indeterminacy of elements in $\mathbb{C}^*$, nor sign indeterminacy). Usually the torsion is an element in the field $\mathbb{C}(t)$, but here it belongs to the ring $\mathbb{C}[t^{\pm 1}]$. By Kitano's theorem [11]

$$\Delta_\rho(t) = \tau(M; \rho \otimes \varphi). \tag{2.1}$$

3. Twisted cohomology of the Seifert surface

Let $S \subset S^3$ be a Seifert surface, namely an embedded orientable and compact surface with boundary $\partial S = K$. Let $\dot{S}$ denote its interior, so that it embeds in the knot complement $M = S^3 \setminus K$. Let's cut off the knot complement along $\dot{S}$:

$$N = (S^3 \setminus K) \setminus \mathcal{N}(\dot{S}),$$

where $\mathcal{N}(\dot{S})$ stands for an open tubular neighborhood of $\dot{S}$. The manifold N is not compact, it can be compactified by adding the knot K:

$$\overline{N} = N \cup K.$$

The boundary of N has two components parallel to $\dot{S}$:

$$\partial N = \dot{S}_- \cup \dot{S}_+.$$

Its compactification is a parallel copy of the Seifert surface $S_\pm = \dot{S}_\pm \cup\ K$. Let us also consider the closed surface

$$F = S_- \cup S_+ \cup K = S_+ \cup_\partial S_-$$

that has twice the genus of S and satisfies $\partial\overline{N} = F$.

From now on ρ denotes a *lift of the holonomy.* So the restriction $\rho|_{\pi_1(S)}$ is a faithful representation and therefore it is irreducible, namely $\mathbb{C}^2$ has no proper invariant subspace. In particular 0 is the only element of $\mathbb{C}^2$ fixed by ρ, so

$$H^0(S;\rho) = 0, \tag{3.1}$$

cf. for instance [2]. Therefore, since S has the homotopy type of a 1-complex and genus g,

$$H^i(S;\rho) = 0, \qquad \text{for } i \neq 1, \tag{3.2}$$

and

$$\dim H^1(S;\rho) = -\chi(S)\cdot \dim \mathbb{C}^2 = 4g-2. \tag{3.3}$$

Then, since $H^*(M;\rho) = 0$, a Mayer-Vietoris argument yields:

$$\dim H^i(N;\rho) = \begin{cases} 0 & \text{for } i \neq 1 \\ 4g-2 & \text{for } i = 1 \end{cases} \tag{3.4}$$

The main result of this section is the following lemma.

Lemma 3.1 *Let $i^*_\pm : H^1(N) \to H^1(\dot{S}_\pm)$ denote the natural maps induced by inclusion. Then*

$$\deg(\Delta_\rho) = \operatorname{rank}(i^*_+) + \operatorname{rank}(i^*_-) - (4g-2).$$

Proof. The proof is based on the Mayer-Vietoris exact sequence applied to the pair $U = N$ and $V = \overline{\mathcal{N}(S)} \cong S \times [-1,1]$, a closed tubular neighborhood of $\dot{S}$ in M, so that $U \cup V = M$ and $U \cap V = S_+ \sqcup S_-$. I shall consider the cohomology and the torsion twisted by $\rho\otimes\varphi$, and in this proof, in order to compute this torsion, I work with the field $\mathbb{C}(t)$, not the ring $\mathbb{C}[t^{\pm 1}]$.

Notice that in each of the spaces U, V and $U \cap V$, the representation ab, and therefore φ, is trivial. Hence $H^*(U;\rho\otimes\varphi) \cong H^*(U;\rho)\otimes\mathbb{C}(t)$, and similarly for V and $U\cap V$. In particular, the Reidemeister torsion $\tau(U;\rho\otimes\varphi)$ is a trivial polynomial, and so is for V and the components of $U\cap V$. By (3.2), (3.3), and (3.4), the only nontrivial map of the long exact sequence in cohomology is

$$(H^1(\dot{S}_+;\rho)\otimes\mathbb{C}(t))\oplus(H^1(\dot{S}_-;\rho)\otimes\mathbb{C}(t)) \to (H^1(N;\rho)\otimes\mathbb{C}(t))\oplus(H^1(\dot{S};\rho)\otimes\mathbb{C}(t)) \tag{3.5}$$

Now, since the Reidemeister torsion is a trivial polynomial for N and $\dot{S}_\pm$, Milnor's formula [16] for the torsion of long exact sequences says that $\tau(M;\rho\otimes\varphi)$ is the determinant of the map (3.5), provided one has a coherent choice of basis in cohomology (which in this case does not matter, as the torsions for N and $\dot{S}_\pm$ are units in $\mathbb{C}[t^{\pm1}]$ and I just care about the degree of the polynomials). After a choice of base points for the fundamental groups, (3.5) can be written as a matrix:

$$\begin{pmatrix} i^*_+ & i^*_- \\ t\cdot & \mathrm{id} \end{pmatrix}, \tag{3.6}$$

where id denotes the identity and $t\cdot$ multiplication by t (i_+^* and id are the identity on the factor $\mathbb{C}(t)$). Therefore

$$\Delta_\rho = \det(i_+^* - t \cdot i_-^*) \tag{3.7}$$

up to multiplication by a unit in $\mathbb{C}[t^{\pm 1}]$. The lemma follows from (3.7) and from $\dim H^1(S;\rho) = 4g - 2$. □

Lemma 3.1 explains the upper bound on the degree of Δ_ρ and why it is maximal for a fibered knot, because then N is the product of $S_\pm$ with an interval and the inclusions $i_\pm$ are homotopy equivalences. The aim of the rest of the paper is to prove that $\operatorname{rank}(i_\pm^*) \geq 2g$.

4. A bilinear product in cohomology

There is a nondegenerate and symmetric bilinear product on $H^1(\partial\overline{N};\rho)$, due to Goldman [8]. Recall that $\partial\overline{N} = F = \partial N \cup K$. This bilinear product is combination of the usual wedge product and a nondegenerate ρ-invariant bilinear product on $\mathbb{C}^2$. More precisely, the determinant induces an antisymmetric bilinear paring in $\mathbb{C}^2$ which is $\mathrm{SL}_2(\mathbb{C})$-invariant

$$\begin{array}{ccc} \mathbb{C}^2 \times \mathbb{C}^2 & \to & \mathbb{C} \\ \begin{pmatrix} a \\ b \end{pmatrix} \otimes \begin{pmatrix} c \\ d \end{pmatrix} & \mapsto & \det \begin{pmatrix} a & c \\ b & d \end{pmatrix} \end{array} .$$

Combined with the antisymmetric cup product in cohomology (i.e. the wedge of differential forms when working with de Rham cohomology) it yields a symmetric bilinear product

$$H^1(F;\rho) \times H^1(F,\rho) \to H^2(F,\mathbb{C}) \cong \mathbb{C}. \tag{4.1}$$

This is a perfect pairing (nondegenerate). In fact, it can be used in general to prove Poincaré duality with twisted coefficients. It will be useful to restrict it to $S_\pm$.

Lemma 4.1 *The inclusion induces an isomorphism*

$$H^1(S,\partial S;\rho) \cong H^1(S;\rho).$$

Proof. Since $\partial S = K$ is represented by the longitude $l \in \pi_1(S) \subset \pi_1(M)$, up to conjugation

$$\rho(l) = -\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

[3, 15]. The minus sign is relevant, as this implies that this matrix has no fixed vectors in $\mathbb{C}^2$ other than zero. Therefore $H^0(S;\rho) = 0$ and, by Euler characteristic, $H^*(S;\rho) = 0$. Then the lemma follows from the long exact sequence of the pair $(S, \partial S)$. □

Lemma 4.1 allows to define again a nondegenerate a bilinear pairing

$$H^1(S;\rho)\times H^1(S;\rho)\cong H^1(S;\rho)\times H^1(S,\partial S;\rho)\to H^2(S,\partial S;\mathbb{C})\cong\mathbb{C}. \tag{4.2}$$

Lemma 4.2 *a) The inclusions $S_\pm\subset F$ induce an isomorphism with the orthogonal sum for the pairing:*

$$H^1(F;\rho)\cong H^1(S_+;\rho)\perp H^1(S_-;\rho).$$

In addition the restriction to each factor is the pairing (4.2)*.*

b) Let $\sigma: H^1(S_\pm;\rho)\to H^1(S_\mp;\rho)$ denote the map induced by the identification of $S_\pm$ with S. If $B_\pm$ denotes the bilinear form on $H^1(S_\pm;\rho)$, then

$$\sigma^* B_\pm = -B_\mp.$$

Proof. For the first assertion, the direct sum is proved using the Mayer-Vietoris exact sequence for the pair (S_+,S_-). Notice that in the proof of Lemma 4.1 it is shown that $H^*(S_+\cap S_-;\rho)=H^*(\partial S;\rho)=0$. Orthogonality of the sum and the fact that the restriction to each factor is the pairing (4.2) follow from naturality of the construction of the product, using for instance de Rham cohomology and wedge product of forms. This proves assertion a).

For assertion b), notice that $\dot S_+$ and $\dot S_-$ are identified with opposite orientation and that $H^2(S,\partial S;\mathbb{C})$ occurs in the construction of the bilinear product (4.2). □

The following result is well known. Combined with the previous orthogonal decomposition it is a key tool for the argument.

Lemma 4.3 *The image of $H^1(\bar N;\rho)\to H^1(F;\rho)$ is an isotropic subspace of maximal dimension (i.e. a Lagrangian subspace).*

Proof. As this is well known, I just sketch the proof. It is based on the fact that in the long exact sequence of the pair the following two maps are dual:

$$H^1(\overline{N};\rho)\xrightarrow{i^*}H^1(F;\rho)\xrightarrow{\theta}H^2(\overline{N},F;\rho).$$

Duality means that there is a similarly defined bilinear pairing

$$H^1(\overline{N};\rho)\times H^2(\overline{N},F;\rho)\to H^3(\overline{N},F;\mathbb{C})$$

satisfying

$$\langle i^*(a),b\rangle_F=\langle a,\theta(b)\rangle_{\bar N},\qquad \forall a\in H^1(\overline{N};\rho),\ b\in H^1(F;\rho),$$

where $\langle,\rangle_F$ and $\langle,\rangle_{\bar N}$ denote the corresponding bilinear products. Then the lemma follows easily form this formula and nondegeneracy. □

5. Lagrangian subspaces

The previous section, specially Lemma 4.3, justifies to work with Lagrangian subspaces of $H^1(F;\rho) \cong \mathbb{C}^{8g-4}$ and also of $H^1(S;\rho) \cong \mathbb{C}^{4g-2}$. Consider $\mathbb{C}^{2n}$ equipped with a nondegenerate symmetric bilinear form. The set of all linear Lagrangian subspaces

$$\mathcal{L}_{2n} = \{L \subset \mathbb{C}^{2n} \mid L \text{ is Lagrangian}\}$$

is an algebraic variety. Let me quote the following result [10, Section 6.1]:

Proposition 5.1 *The variety $\mathcal{L}_{2n}$ has two irreducible components. In addition, given L_0 and L_1 in $\mathcal{L}_{2n}$,*

$$\dim(L_0 \cap L_1) \cong n \mod 2$$

if and only if L_0 and L_1 belong to the same component.

Notice that this condition on the dimension implies that the irreducible components of $\mathcal{L}_{2n}$ are disjoint.

Later I will apply Proposition 5.1 to $H^1(F;\rho) \cong \mathbb{C}^{8g-4}$. Consider in addition an orthogonal decomposition

$$\mathbb{C}^{8g-4} \cong \mathbb{C}^{4g-2} \perp \mathbb{C}^{4g-2}.$$

In view of Lemma 4.2, assume that the bilinear form decomposes as

$$B \oplus -B.$$

Let $\sigma : \mathbb{C}^{4g-2} \to \mathbb{C}^{4g-2}$ denote the permutation of factors in the orthogonal sum. Set

$$\begin{aligned} S^{eq} =& \{L_1 \oplus L_2 \in \mathcal{L}_{8g-4} \mid L_1 \text{ and } \sigma(L_2) \text{ belong to } \textit{the same} \text{ component of } \mathcal{L}_{4g-2}\}, \\ S^{dif} =& \{L_1 \oplus L_2 \in \mathcal{L}_{8g-4} \mid L_1 \text{ and } \sigma(L_2) \text{ belong to } \textit{different} \text{ components of } \mathcal{L}_{4g-2}\}. \end{aligned}$$

Corollary 5.1 *The sets S^{eq} and S^{dif} belong to different components of $\mathcal{L}_{8g-4}$.*

Proof. The proof is straightforward from the formula

$$\dim((L_1 \perp L_2) \cap (L_1' \perp L_2')) = \dim(L_1 \cap L_1') + \dim(L_2 \cap L_2')$$

for $L_i, L_i' \in \mathcal{L}_{4g-2}$, and from Proposition 5.1. □

The corresponding components of $\mathcal{L}_{8g-4}$ are denoted by $\mathcal{L}^{eq}$ and $\mathcal{L}^{dif}$. Let

$$\pi_\pm : H^1(S_+;\rho) \oplus H^1(S_-;\rho) \to H^1(S_\pm;\rho)$$

denote the projection to each factor.

Lemma 5.1 *Let $L \in \mathcal{L}_{8g-4}$ be a Lagrangian subspace such that $\pi_+(L) = H^1(S_+;\rho)$ (or $\pi_-(L) = H^1(S_-;\rho)$). Then $L \in \mathcal{L}^{eq}$.*

Proof. The restriction of π_+ defines an isomorphism $L \cong H^1(S_+;\rho)$. Choose $\mathcal{B} = \{e_1, \dots, e_{4g-2}\}$ a basis for $H^1(S_+;\rho)$. Then $\mathcal{B}' = \sigma(\pi_-(\pi_+(\mathcal{B})))$ is another basis. Since L is Lagrangian and the linear form is $B_+ \perp B_-$, with $\sigma^* B_- = -B_+$, one has that $\mathcal{B}' = A(\mathcal{B})$ for some $A \in O(4g-2, \mathbb{C})$. In fact, one can always find $A \in SO(4g-2, \mathbb{C})$ so that $e_i' = A(e_i)$ for $i = 1, \dots, 4g-3$ and $e'_{4g-2} = \pm A(e_{4g-2})$. Since $B_+(e_i, e_j) = -B_-(e_i', e_j')$, the sign indeterminacy can be removed: $e'_{4g-2} = A(e_{4g-2})$.

As $SO(4g-2, \mathbb{C})$ is connected, A can be deformed in $SO(4g-2, \mathbb{C})$ to the identity matrix, hence $L \in \mathcal{L}^{eq}$. $\square$

6. Proof of Theorem 1.2

By Lemma 3.1 it suffices to show that $\operatorname{rank}(i_\pm^*) \geq 2g$, where $i_\pm^* : H^1(N;\rho) \to H^1(S_\pm;\rho)$ denote the maps induced by inclusion.

Lemma 6.1 $\dim \ker i_\pm^* \leq 2g-1$.

Proof. The image $i_\mp^*(\ker i_\pm^*)$ is an isotropic subspace of $H^1(S_\mp;\rho) \subset H^1(F;\rho)$. In addition, $i_\mp^*$ is injective on $\ker(i_\pm^*)$, as $(i_+^*, i_-^*) : H^1(\bar N;\rho) \to H^1(F;\rho)$ is injective. Therefore $\dim(\ker i_\pm^*) = \dim(i_\mp^*(\ker i_\pm^*)) \leq \frac{1}{2} \dim H^1(S_\mp;\rho) = 2g-1$. $\square$

This lemma implies that $\operatorname{rank}(i_\pm^*) \geq 2g-1$. So one may assume that $\operatorname{rank}(i_+^*) = 2g-1$ or $\operatorname{rank}(i_-^*) = 2g-1$ and look for a contradiction.

Lemma 6.2 *If* $\operatorname{rank}(i_+^*) = 2g-1$ *then* $\operatorname{rank}(i_-^*) = 2g-1 = \dim\ker(i_\pm^*)$.

Proof. $\dim\ker(i_+^*) = \dim H^1(N;\rho) - \operatorname{rank}(i_+^*) = 2g-1$. Chose a complement, namely a subspace $V \subset H^1(N;\rho)$ so that $H^1(N;\rho) = \ker(i_+^*) \oplus V$. Let $i : F \subset \bar N$, denote the inclusion. Since $i^* = i_+^* \perp i_-^*$ and the image of i^* is isotropic, then $i^*(V) \perp i^*(\ker(i_+^*))$. As $i^*(\ker(i_+^*)) = i_-^*(\ker(i_+^*))$ is a Lagrangian subspace of $H^1(S_-;\rho)$, it follows that $i_-^*(V) \subset i_-^*(\ker(i_+^*))$. Namely $i_-^*(H^1(N;\rho)) = i_-^*(\ker(i_+^*))$ and $\operatorname{rank}(i_-^*) = 2g-1$. $\square$

Since $H^*(M;\rho) = 0$, a Mayer-Vietoris argument yields

$$\sigma(i_-^*(H^1(N;\rho))) \cap i_+^*(H^1(N;\rho)) = 0,$$

here σ is the map induced by the identification of S_+ with S_-. Thus, by Proposition 5.1 and as $2g-1$ is odd, $\sigma(i_-^*(H^1(N;\rho)))$ and $i_+^*(H^1(N;\rho))$ belong to different components of $\mathcal{L}_{4g-2}$. By Corollary 5.1, $i_*(H^1(N;\rho)) \in \mathcal{L}^{dif}$.

I'll find a contradiction by showing that $i_*(H^1(N;\rho)) \in \mathcal{L}^{eq}$. The manifold M can be viewed as the result of gluing N and $P = \dot{S} \times [-1,1]$ along the boundary. Notice that the inclusion induces an isomorphism $H^1(P;\rho) \cong H^1(\dot{S} \times \{\pm 1\};\rho)$ for each component of the boundary $\dot{S} \times \{\pm 1\}$. In particular Lemma 5.1 applies, hence the image of $H^1(P;\rho) \cong H^1(\dot{S} \times \{-1\};\rho) \perp H^1(\dot{S} \times \{+1\};\rho)$ lies in $\mathcal{L}^{eq}$. Since the intersection of the image of $H^1(P;\rho)$ with the image of $H^1(N;\rho)$ is trivial and $4g-2$ is even, by Proposition 5.1 they lie in the same component, which is $\mathcal{L}^{eq}$.

References

[1] I. Agol, N. M. Dunfield: *Certifying the thurston norm via SL(2, C)-twisted homology.* arXiv:1501.02136, 2015.

[2] K. S. Brown: *Cohomology of groups.* GTM 87, Springer-Verlag, New York 1994 (Corrected reprint of the 1982 original).

[3] D. Calegari: *Real places and torus bundles. Geom. Dedicata* 118 (2006) 209–227.

[4] M. Culler: *Lifting representations to covering groups.* Adv. in Math. 59 (1) (1986) 64–70.

[5] N. M. Dunfield: *Alexander and Thurston norms of fibered 3-manifolds.* Pacific J. Math. 200 (1) (2001) 43–58.

[6] N. M. Dunfield, S. Friedl, N. Jackson: *Twisted Alexander polynomials of hyperbolic knots.* Exp. Math. 21 (4) (2012) 329–352.

[7] S. Friedl, S. Vidussi: *A survey of twisted Alexander polynomials.* In *The mathematics of knots,* Contrib. Math. Comput. Sci. vol.1, 45–94. Springer, Heidelberg 2011.

[8] W. M. Goldman: *The symplectic nature of fundamental groups of surfaces.* Adv. in Math. 54 (2) (1984) 200–225.

[9] F. González-Acuña, José María Montesinos-Amilibia: *On the character variety of group representations in* $\mathrm{SL}(2,\mathbf{C})$ *and* $\mathrm{PSL}(2,\mathbf{C})$. Math. Z. 214 (4) (1993) 627–652.

[10] P. Griffiths, J. Harris: *Principles of algebraic geometry.* Pure and Applied Mathematics, Wiley-Interscience [John Wiley & Sons], New York 1978.

[11] T. Kitano: *Twisted Alexander polynomial and Reidemeister torsion.* Pacific J. Math. 174 (2) (1996) 431–442.

[12] X.-S. Lin: *Representations of knot groups and twisted Alexander polynomials.* Acta Math. Sin. (Engl. Ser.) 17 (3) (2001) 361–380.

[13] C. T. McMullen: *The Alexander polynomial of a 3-manifold and the Thurston norm on cohomology.* Ann. Sci. École Norm. Sup. (4) 35 (2) (2002) 153–171.

[14] P. Menal-Ferrer, J. Porti: *Mutation and* $\mathrm{SL}(2,\mathbb{C})$*-Reidemeister torsion for hyperbolic knots.* Algebr. Geom. Topol. 12 (4) (2012) 2049–2067.

[15] P. Menal-Ferrer, J. Porti: *Twisted cohomology for hyperbolic three manifolds. Osaka J. Math.*, 49(3):741–769, 2012.

[16] J. Milnor: *Whitehead torsion.* Bull. AMS 72 (1966) 358–426.

[17] J. Milnor: *Infinite cyclic coverings.* In *Conference on the Topology of Manifolds* (Michigan State Univ., East Lansing, Mich.) 115–133, 1968.

[18] J. Milnor: *A duality theorem for Reidemeister torsion.* Ann. of Math. (2) 76 (1962) 137–147.

[19] W. P. Thurston *The Geometry and Topology of Three-Manifolds.* Notes, 1980 (Electronic 2002 version at http://www.msri.org/publications/books/gt3m/).

[20] V. G. Turaev: *Reidemeister torsion in knot theory.* Uspekhi Mat. Nauk 41 (1(247)) (1986) 97–147, 240.

[21] V. G. Turaev: *Introduction to combinatorial torsions.* Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel 2001. Notes taken by Felix Schlenk.

[22] M. Wada: *Twisted Alexander polynomial for finitely presentable groups.* Topology 33 (2) (1994) 241–256.

Matrices de rotaciones, simetrías y roto-simetrías

María Jesús de la Puente

Departamento de Álgebra
Facultad de Matemáticas
Universidad Complutense
28040 Madrid, Spain
mpuente@ucm.es

Resumen

In this note we find the orthogonal matrices $R, S \in M_3(\mathbb{R})$ corresponding to the clockwise rotation r in $\mathbb{R}^3$ around the axis generated by a unit vector $u = (a, b, c)^t$ through an angle $\alpha \in [0, 2\pi)$, and to the symmetry s in $\mathbb{R}^3$ on the plane perpendicular to u. Matrix S depends on a, b, c and matrix R depends on $a, b, c, \cos\alpha$ and $\sin\alpha$. We show $SR = RS$.

2010 Mathematics Subject Classification: 15B10.
Key words: matriz ortogonal; rotación; simetría; proyección

En los libros de Álgebra Lineal al uso a nivel universitario, se encuentran diversos problemas en el espacio euclídeo $\mathbb{R}^3$ del siguiente tipo:

1. dados la recta E y el ángulo $\alpha \in [0, 2\pi)$, hallar la matriz R de la rotación (o giro) en $\mathbb{R}^3$ alrededor del eje E, de amplitud α y sentido positivo,

2. dado el plano H, hallar la matriz S de la simetría (o reflexión) en $\mathbb{R}^3$ sobre H,

donde E, α y H son datos concretos. Llamaremos a estos *problemas directos*. Bajo hipótesis adecuadas, las matrices R, S resultan ser ortogonales, i.e., $R^{-1} = R^t$ y $S^{-1} = S^t$. Asimismo encontramos los *problemas inversos*: dada la matriz ortogonal $M \in M_3(\mathbb{R})$, averiguar si M representa una rotación alrededor de un eje, o una simetría, o la composición de las anteriores, indicando en cada caso los elementos geométricos asociados (eje E, amplitud $\alpha \in [0, 2\pi)$ de la rotación, plano H de la simetría, etc.)[1]

[1]Podemos hablar también de rotación de amplitud $\alpha \in [0, \pi]$ con sentido positivo o negativo. La equivalencia es clara: la rotación de amplitud α y sentido negativo coincide con la rotación de amplitud $2\pi - \alpha$ y sentido positivo.

En esta nota abordamos las preguntas anteriores en general, lo que sirve para responder tanto problemas directos como inversos. Obtendremos $S \in M_3(F_1)$ y $R \in M_3(F_2)$, para cierta extensión de cuerpos

$$\mathbb{Q} \subseteq F \subseteq F_1 \subseteq F_2 \subseteq \mathbb{R},$$

donde F es el cuerpo base del problema.

Trabajaremos en un espacio vectorial real V de dimensión 3, dotado de un producto escalar $\langle\,,\rangle$ y usaremos siempre bases ortonormales. Las coordenadas de los vectores de V (respecto de cualquier base) se escribirán en columna. Para fijar ideas, el lector puede tomar $V = \mathbb{R}^3$ con el producto escalar usual $\langle\,,\rangle$, el producto vectorial usual $\wedge$ y la base canónica $\{e_1, e_2, e_3\}$. Recordemos la conocida igualdad

$$u \wedge (v \wedge w) = v\langle u, w\rangle - w\langle u, v\rangle, \tag{1}$$

cuya demostración (usando coordenadas) es un sencillo ejercicio.

Fijemos una base ortonormal $\mathcal{B}$ de V. Sean $(a, b, c)^t$ las coordenadas, respecto de $\mathcal{B}$, de un vector unitario u que genere E (i.e., $\langle u, u\rangle = a^2 + b^2 + c^2 = 1$). Denotemos por p_E la proyección ortogonal de $\mathbb{R}^3$ sobre E. Denotemos por $r_{E,\alpha}$ y s_H la rotación y simetría descritas más arriba, donde el plano H, de ecuación $ax + by + cz = 0$, es precisamente $E^\perp$.

Lema. La matriz de p_E respecto de $\mathcal{B}$ es

$$A = \begin{pmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{pmatrix} \in M_3(\mathbb{Q}(a, b, c)). \tag{2}$$

Demostración. Si $v \in V$ es un vector arbitrario, sabemos que $p_E(v) = u\langle u, v\rangle$, ya que u es unitario. Por tanto, si las coordenadas de v respecto de $\mathcal{B}$ son $(x, y, z)^t$, tenemos

$$p_E(v) = \begin{pmatrix} a \\ b \\ c \end{pmatrix} (a, b, c) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = A \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

$\square$

Teorema. La matriz de $r_{E,\alpha}$ es

$$R = I + (\operatorname{sen}\alpha)B + (\cos\alpha - 1)(I - A) \in M_3(\mathbb{Q}(a, b, c, \cos\alpha, \operatorname{sen}\alpha)), \tag{3}$$

y la matriz de $s_{E^\perp}$ es

$$S = I - 2A \in M_3(\mathbb{Q}(a, b, c)), \tag{4}$$

donde I denota la matriz identidad de orden 3 y

$$B = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix}.$$

Demostración. La demostración de (3) requiere varios pasos. Para empezar, observemos que la aplicación $g_u : V \to V$ tal que $v \mapsto u \wedge v$ es lineal. Además, la matriz de g_u respecto de $\mathcal{B}$ es B. Comprobamos que $-B^2 = I - A$, lo que significa que

$$-g_u^2 = \mathrm{id} - p_E = p_{E^\perp} \tag{5}$$

ya que las proyecciones sobre E y sobre $E^\perp$ son complementarias. A continuación veamos que

$$r_{E,\alpha} = \mathrm{id} + (\operatorname{sen}\alpha) g_u + (\cos\alpha - 1) p_{E^\perp}, \tag{6}$$

de donde se deducirá (3), gracias a (5) y al lema. Escribamos $f = \mathrm{id} + (\operatorname{sen}\alpha) g_u + (\cos\alpha - 1) p_{E^\perp}$ y demostremos que f y $r_{E,\alpha}$ actúan igual sobre los elementos de cierta base $\mathcal{B}'$ de V. Tomemos cualquier vector unitario v perpendicular a u y sea $\mathcal{B}' = \{u, v, u \wedge v\}$. Unos cálculos sencillos (usando $u \wedge (u \wedge v) = u\langle u, v\rangle - v\langle u, u\rangle = -v$ a partir de (1)) nos muestran que $f(u) = u$, $f(v) = (\cos\alpha)v + (\operatorname{sen}\alpha)(u \wedge v)$ y $f(u \wedge v) = -(\operatorname{sen}\alpha)v + (\cos\alpha)(u \wedge v)$, de donde se sigue la igualdad (6) y, con ella (3).

Ahora tomemos el vector $v = h^{-1}(-b, a, 0)^t$ y consideramos la base $\mathcal{B}' = \{u, v, u \wedge v\}$, donde $h = \sqrt{a^2 + b^2}$. La matriz de $s_{E^\perp}$ respecto de $\mathcal{B}'$ es $D = \mathrm{diag}(-1, 1, 1)$, ya que v y $u \wedge v$ son perpendiculares a u. Por tanto, la matriz de $s_{E^\perp}$ respecto de $\mathcal{B}'$ es $S = PDP^{-1}$, donde

$$P = \begin{pmatrix} a & -b/h & -ac/h \\ b & a/h & -bc/h \\ c & 0 & h \end{pmatrix}$$

es matriz ortogonal (i.e., $P^{-1} = P^t$). Un cálculo sencillo proporciona $S = PDP^t = I - 2A$, que es la expresión (4). $\square$

Corolario. Llamemos *roto-simetría* a la composición $s_{E^\perp} \circ r_{E,\alpha} = r_{E,\alpha} \circ s_{E^\perp}$. Su matriz es

$$SR = RS = S + (\operatorname{sen}\alpha)B + (\cos\alpha - 1)(I - A) \in M_s(\mathbb{Q}(a, b, c, \cos\alpha, \operatorname{sen}\alpha)). \tag{7}$$

Demostración. Basta ver que rotación y simetría conmutan, y esto es cierto ya que $SB = B = BS$ y $S(I - A) = I - A = (I - A)S$, igualdades de comprobación inmediata. $\square$

Observaciones.

1. En (4) tenemos $S = I - 2A$, de donde se deduce la conocida relación (ver figura 1)
$$s_{E^\perp} = \mathrm{id} - 2p_E. \tag{8}$$

2. El rango de la matriz A es 1 y A no es ortogonal. Se verifica $A^2 = A = A^t$ e $(I - A)^2 = I - A$. La matriz B es antisimétrica y $-B^2 = I - A$. De aquí se sigue que las matrices R y S son ortogonales, i.e., $RR^t = I = SS^t = S^2$.

3. Los determinantes y las trazas valen $\det R = 1$, $\det S = \det(SR) = -1$, $\mathrm{tr}R = 1 + 2\cos\alpha$, $\mathrm{tr}S = 1$ y $\mathrm{tr}(SR) = -1 + 2\cos\alpha$. Determinante y traza son valores invariantes de una isometría de $\mathbb{R}^3$, pero NO la caracterizan en general (salvo que la traza valga 1). En efecto, el determinante nos dice si la isometría conserva o invierte la orientación y la traza proporciona el valor $\cos\alpha$. Queda, pues, por determinar el signo de $\operatorname{sen}\alpha = \pm\sqrt{1 - \cos^2\alpha}$.

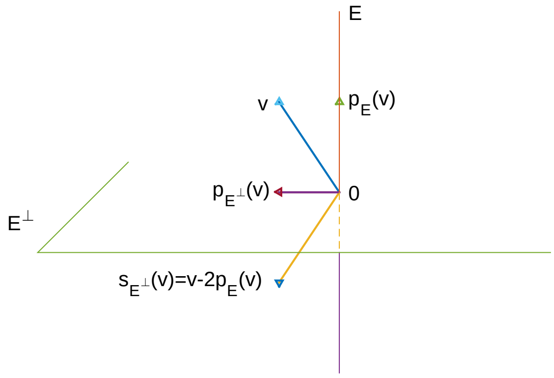


Figura 1: Simetría respecto del plano perpendicular a la recta E.

Ejemplo. Una matriz ortogonal sencilla (pero no trivial) es

$$M = \frac{1}{pq}\begin{pmatrix} p & q & 1 \\ p & -q & 1 \\ p & 0 & -2 \end{pmatrix} \in M_3(F),$$

donde $p = \sqrt{2}$ y $q = \sqrt{3}$ y $F = \mathbb{Q}(p, q)$ es el cuerpo base. Como $MM^t = I$ y $\det M = 1$, sabemos que M representa una rotación alrededor de un eje E. Vamos a hallar un vector unitario $u = (a, b, c)^t$ que genere E y la amplitud de giro $\alpha \in [0, 2\pi)$ en sentido positivo. Ciertos cálculos proporcionan

$$a = \frac{p+q}{n}, \quad b = \frac{2 - p - q + pq}{n}, \quad c = \frac{1}{n},$$

con $n = \sqrt{21 - 10p - 8q + 8pq}$. Como la traza es invariante sabemos, por la observación 3, que

$$1 + 2\cos\alpha = \mathrm{tr}M = \frac{-2 + p - q}{pq} = -\frac{p}{2} + \frac{q}{3} - \frac{pq}{3},$$

de donde

$$\cos\alpha = -\frac{1}{2} - \frac{p}{4} + \frac{q}{6} - \frac{pq}{6}, \quad \operatorname{sen}\alpha = \pm\sqrt{1 - \cos^2\alpha} = \pm\frac{pqr}{12},$$

con $r = \sqrt{9 - 2p - 2pq}$. Sustituimos estos valores en (3) y obtenemos la igualdad $M = R$ cuando el signo del seno es NEGATIVO, i.e., $\operatorname{sen}\alpha = -pqr/12$. Conocidos $\operatorname{sen}\alpha$ y $\cos\alpha$, deducimos que $\alpha \in (\pi, 2\pi/3)$; concretamente $\alpha \simeq 193°20'$.

Los libros de texto suelen usar otro procedimiento para calcular α. Hallan la amplitud

$$\alpha = \arccos\left(-\frac{1}{2} - \frac{p}{4} + \frac{q}{6} - \frac{pq}{6}\right),$$

que tiene DOS soluciones, $\alpha_1, \alpha_2 \in [0, 2\pi)$, con $\alpha_1 \leq \alpha_2 = 2\pi - \alpha_1$. Luego determinan, mediante algún argumento geométrico, si $\alpha = \alpha_1$ ó $\alpha = \alpha_2$. En cambio, nuestro razonamiento se ha basado en las fórmulas (3), (4) y (7), trabajando en la extensión algebraica $\mathbb{Q}(p, q, n, r)$ del cuerpo base $\mathbb{Q}(p, q)$. En nuestro caso tenemos $\cos\alpha \simeq -0{,}973126$, $\operatorname{sen}\alpha \simeq -0{,}230270$ y $\alpha = \alpha_2 \simeq 193°20'$.

Agradecimientos. He tomado la matriz R (así como su obtención) de la parte debida a Roger C. Alperin en el libro [1], p. 113–114. Recomiendo vivamente este texto a todos los profesores universitarios de Algebra Lineal: en el encontrarán verdaderas joyas.

Referencias

[1] D. Carlson et al. (eds.): *Linear algebra gems. Assets for undergraduate mathematics.* MAA Notes series **59,** 2002.

La Biblioteca de El Escorial.
Un culto a la matemática

María Concepción Romo Santos

Departamento de Álgebra
Universidad Complutense de Madrid
28040 Madrid, Spain
romosan@mat.ucm.es

Dedicado al profesor José María Montesinos Amilibia con ocasión de su jubilación académica, como homenaje a su persona y en agradecimiento por sus enseñanzas y ejemplo constante.

Resumen

Philip II and his architect Juan de Herrera had the idea of perpetuating in El Escorial a special worship to the Mathematics and test of it is the construction of the Library of the Monastery.

2010 Mathematics Subject Classification: 01
Key words: Mathematics, astronomy and astrolabe.

1. Introducción

El Monasterio de San Lorenzo de El Escorial es hoy eje geográfico y espiritual de España, convertido en polo de peregrinación para todos los interesados en calar las esencias de nuestra historia y de nuestra cultura. No se puede estudiar la edad de Oro española sin que la presencia de El Escorial irradie desde el horizonte, con los guiños de una España que pretendió ser eterna y que casi lo consiguió. El Escorial es la expresión de la razón geométrica y del estudio minucioso de todos los detalles arquitectónicos, artísticos y litúrgicos. Felipe II y su arquitecto Juan de Herrera tuvieron la idea de perpetuar en El Escorial un culto especial a la matemática y prueba de ello es la construcción de la Biblioteca.

2. Felipe II, defensor e impulsor de las matemáticas

El rey Felipe II fue discípulo de Juan Martínez Silíceo, uno de los matemáticos más prestigiosos de la España de la primera mitad del siglo XVI. En cuanto a sus ideas filosóficas, fue un ardiente defensor y divulgador de las teorías de Raimon Llull. Durante su reinado, y en lo concerniente a la ciencia, fue un personaje abierto a todo el saber científico y un gran defensor del estudio de las Matemáticas y prueba de ello es que decidió instituir una Academia de Matemáticas en Madrid.


Felipe II y Juan de Herrera

Juan de Herrera, el gran colaborador de Felipe II, nació en Mobellán, estudió en la Universidad de Valladolid y siguiendo a Carlos I recorrió Flandes e Italia. En 1.556 volvió a España acompañando a Carlos I en su retiro a Yuste, hasta su muerte en 1.558. En este momento Juan de Herrera se incorpora a la corte de Felipe II y es enviado a Alcalá de Henares para diseñar las figuras geométricas del Libro de los Saberes, de Alfonso X el Sabio. En 1.562 entra al servicio de Juan Bautista de Toledo, director de la construcción del Monasterio de El Escorial.

En 1.567 muere Juan Bautista de Toledo y en virtud de la fama adquirida como hombre esmerado en la realización de figuras geométricas, se le encarga la dirección en la continuación de las obras de El Escorial. Los descubrimientos de Juan de Herrera de máquinas y grúas, suponen una ayuda inestimable para la construcción de tan impresionante obra.

El desarrollo de su actividad como arquitecto, viene marcado por la Geometría. Toda su obra es de traza geométrica. Sus contemporáneos le estimaron en gran manera y dijeron de él que era un gran matemático.

3. La Biblioteca de El Escorial

La prueba concluyente de que en la concepción por Felipe II, y en la ejecución por su arquitecto Juan de Herrera, preside la idea de perpetuar en El Escorial un culto especial a la Matemática, como parte integrante de la cultura científica, es la Biblioteca. Haciendo un análisis de la ornamentación, resulta evidente la correlación de la simbología con esta idea. En la amplia bóveda, pueden observarse cuatro gigantescas matronas, que simbolizan la Aritmética, la Música, la Geometría y la Astronomía. En las ramas laterales se encuentran las figuras de Terentino, Pitágoras, Jenofonte, Arquímedes, Jordano Nemorario, Sacrobosco y Regiomontano, con algún que otro músico y los astrónomos, tan relacionados con la Geometría, Ptolomeo y Alfonso X el Sabio. En los paños inferiores se representan escenas históricas, plenas de simbolismo matemático: Salomón proponiendo enigmas ante un peso de balanza, una regla y un ábaco. Arquímedes absorto en una demostración geométrica, se desentiende de los soldados romanos que le amenazan de muerte. Toda la ornamentación es, como se puede observar, un homenaje a la Matemática. La Biblioteca del Real Monasterio de El Escorial es una de las mejores de Europa, y sin duda sería la mejor del mundo si, en 1.671, un incendio no hubiese destruido más de 4.000 manuscritos; si, en 1.809, no hubieran entrado las tropas de Napoleón, o si no hubieran desaparecido miles de volúmenes durante el Trienio Liberal (1.820–23). A pesar de todo, la Biblioteca de El Escorial cuenta con 60.000 volúmenes y cerca de 5.000 manuscritos. Felipe II, el gran creador de esta biblioteca, fue un ardiente defensor y divulgador de todo el saber científico de su época, prueba de ello es la adquisición de las mejores bibliotecas de su tiempo con el fin de surtir a la librería real del Monasterio de El Escorial. Las tres principales bibliotecas que adquirió Felipe II fueron: La Biblioteca de Muley Zadam, la Biblioteca de Juan de Herrera y la Biblioteca de Paez de Castro.

La Biblioteca que Felipe II funda en El Escorial es la primera renacentista que se instala en la nación, acorde con las ideas modernistas que sobre estas instituciones tenía el monarca, adquiridas tanto en las visitas de librerías que vio en sus viajes por Italia, Alemania, Inglaterra y Países Bajos, como por las sugerencias de sus asesores, expertos asistentes a Bibliotecas extranjeras, en especial el doctor Paéz de Castro, que tantos años había pasado manejando bibliotecas en Venecia y Roma. Con el fin de probar la afirmación anterior seguiremos a Teodoro Martín en su libro: "Vida y obra de Juan Páez de Castro", editado por la Diputación Provincial de Guadalajara. En él se recoge un memorial que dirigió Páez de Castro a Felipe II sobre la importancia de establecer librerías (Bibliotecas) reales en el reino: Precisamente al seguir D. Felipe el consejo de Páez, se decidió a fundar una de las bibliotecas más selectas de Europa. El memorial entre otras cosas decía: « *Tras los libros van los hombres sabios y tras ellos los* *que quieren ser sus discípulos. Por causa de las Librerías perdieron muchas Naciones el nombre de bárbaras, y muchas Ciudades fueron frecuentadas de los principales Hombres del Mundo, y se ennoblecieron con Estudios y Universidades. Las librerías son causa, que se haga amistad, y concordia entre muy diversas Naciones por vía de letras. No creciera tanto Alexandría, si, aquella librería no atraxera tantos Sabios, que hicieron aquella tan famosa Universidad».*

4. Principales libros de astronomía en la Biblioteca de El Escorial

La ciencia de los astros obtuvo un notable desarrollo en el siglo XVI, con el apoyo y protección de Felipe II. El recuerdo de algunos hombres célebres de ésta época y sus obras científicas confirman el progreso de los estudios del mundo celeste.

Andrés García Céspedes escribió “Las teorías de los planetas” y propuso a Felipe II la fundación de un gran Observatorio astronómico en El Escorial, comprometiéndose a construir todos los instrumentos que presentó para este proyecto. Rodrigo Zamorano calculó con gran precisión los 33 eclipses que habían de tener lugar desde 1.584 hasta 1.605, publicando láminas de ellos. Nebrija publicó un tratado de Cosmografía, dando nueva forma a los estudios astronómicos. El propio Felipe II dio instrucciones a Juan López de Velasco para observar el eclipse de sol de 26 de febrero de 1.577, con instrumentos de su invención.

Un fiel reflejo del progreso de la Astronomía de la época de Felipe II es la excelente colección bibliográfica de temas astronómicos y astrológicos de la Real Biblioteca de El Escorial. Citaremos los autores y textos más significativos.

Juan Sacrobosco, cuyo tratado de la esfera hizo universalmente célebre el Almagesto, único manual de enseñanza hasta que en el siglo XV Peurbachio dio a conocer su “Teórica Novae Planetarium”, está presente en la Biblioteca en cantidad de manuscritos e impresos. Raimond Llul es fundamentalmente un filósofo, pero tenía una vastísima erudición enciclopédica, dedicándose también a la Astronomía y a la Matemática. En la Biblioteca Escurialense encontramos, además de la “Ars Magna”, varios tratados de Astrología y Astronomía, en forma manuscrita, fechados en París, 1.927.

Jorge Peurbachio se encuentra ampliamente representado en la Biblioteca Escurialense. Su “Teórica Novae Planetarium” representa una vuelta a los árabes, en esta obra, desarrolla las teorías astronómicas del Sol, la Luna y los Planetas, siguiendo las ideas de Tabit ben Qurra.

Juan Regiomontano es el gran astrónomo y el gran divulgador de la matemática del siglo XV. Sus tablas trigonométricas fueron muy utilizadas a lo largo de los siglos XV y XVI, y como divulgador dio a conocer, además del Almagesto, la Música, la Óptica y la Geometría de Ptolomeo, las Sphericas de Teodosio, las Cónicas de Apolonio y las Mecánicas de Aristóteles y Herón. En la Biblioteca de EL Escorial encontramos una edición del Almagesto de Venecia 1.496.

El siglo XVI; que comienza con la revolución matemática desencadenada por la Summa, da luz, también a una revolución más profunda aún, con la aparición del “Revolutionibus” de Copérnico, que aporta una concepción nueva para el sistema del universo. El sistema ptolemaico es reemplazado por el sistema heliocéntrico que supone una gran convulsión de la concepción científica. Del “Revolutionibus” se encuentra, en la Biblioteca de El Escorial, una edición de Nuremberg 1.543 y otra de su discípulo Raethicus, de Basilea 1.566.

Una figura interesante, que encontramos en la Biblioteca Escurialense, es el cosmógrafo alemán Pedro Apiano, con una gran cantidad de volúmenes, entre los que destacaremos la "Cosmographucus" en una edición de Amberes 1.533, en la que explica el Cuadrante Universal, dando una tabla de senos, de minuto en minuto, que fue utilizada mucho a lo largo del siglo. La obra de Cristophoro Clavio tiene la importancia de contribuir grandemente a la propagación de las Matemáticas, a través de los colegios y universidades de la Compañía de Jesús, a la que pertenecía y especialmente por sus obras de las que Descartes aprendió y gustó de las Matemáticas. En la Biblioteca de El Escorial se encuentran muchas de sus obras, entre otras: "Novi Calendarii Romani", realizado en 1.588 por encargo del Papa Gregorio XIII. "El Astrolabium", Roma 1.593, en el que hace unos estudios de la Trigonometría, incluyendo la fabricación y uso del astrolabio.

Después de este recorrido por los textos de la biblioteca, comprobamos con toda seguridad la siguiente afirmación: «La Astronomía fue una de las ciencias que mayores beneficios recibió del rey Felipe II».

Referencias

[1] T. Martín Martín: *Vida y obra de Juan Páez de Castro.* Instituto Provincial de Cultura Marqués de Santillana de Guadalajara, 1990.

[2] P. Vicente, J. Andrés: *Presencia de las Matemáticas en el Monasterio de El Escorial.* Seminario de Historia de la Matemática I. UCM 1991.

[3] Mª C. Romo Santos: *El Escorial, imperio de la Geometría y del Número.* Revista de la Academia Canaria de Ciencias **X** (1998).

[4] J. A. Sánchez Pérez: *Las Matemáticas en la Biblioteca de El Escorial.* Publicaciones del Instituto de Estudios Árabes del Consejo Superior de Investigaciones Científicas. Madrid, 1929.

Geometría en el siglo XIV: los trabajos de Thomas Bradwardine

Juan Tarrés Freixenet

Departmento de Geometría y Topología
Universidad Complutense de Madrid
28040 Madrid, España

A mi querido amigo José María Montesinos como testimonio de una amistad de hace ya muchos años.

1. Introducción

Durante la Edad Media, el estudio de la Geometría en el occidente europeo estaba incluido en lo que se llamó el *Quadrivium*, heredero de los *Mathemata* pitagóricos y que estaba constituido por los estudios de Aritmética, Geometría, Astronomía y Música. Más tarde, se incorporó al sistema lo que llamaban el *Trivium*, formado por la Gramática, la Dialéctica y la Retórica, configurando lo que se conoce como *Las siete artes liberales.*

Las siete artes liberales se mencionan ya a principios del siglo V. En la segunda mitad del siglo VI, Casiodoro intentó cristianizarlas y sistematizarlas como un cuerpo enciclopédico de conocimientos en sus *Institutiones saecularum literarum.* Su uso en las escuelas monásticas y catedralicias de la Alta Edad Media generalizaron el concepto, que se fijó a finales del siglo VIII cuando se adoptaron como currículo educativo por Alcuino de York (735-804) para la Escuela Palatina de Aquisgrán.

La implantación de las siete disciplinas se generalizó primero en el Imperio Carolingio para acabar extendiéndose en todos los reinos de Europa occidental. En las Universidades, al trabajo preparatorio del *trivium* le seguían las enseñanzas superiores del *quadrivium*, esquema que ha pasado a conocerse como "educación clásica" y que no sufrió innovaciones hasta la llegada del denominado renacimiento del siglo XII.

Los textos que se utilizaban en las enseñanzas del *quadrivium* estaban prácticamente todos basados en las obras que Boecio, patricio romano que vivió entre los años 480 y 524, había escrito para tales fines, libros que no eran sino simples resúmenes a un nivel muy elemental de las obras de los matemáticos griegos. Escribió una *Aritmética* basada en la *Introductio Arithmethicae* de Nicómaco de Gerasa, una *Geometría* que contiene solamente proposiciones, sin demostración, de algunas de las partes más sencillas de los cuatro primeros libros de los Elementos de Euclides, una *Astronomía*, resumen del *Almagesto* de Ptolomeo, y finalmente una *Música*, basada en obras anteriores de Euclides, Nicómaco y Ptolomeo. Estas obras fueron muy utilizadas en las escuelas de los monasterios medievales, bien en su forma original, bien, más frecuentemente, con interpolaciones y variantes. En ellas, de la estructura rigurosa y formal de las matemáticas griegas, apenas queda ningún rastro.

A partir del siglo XII se recuperan algunos textos griegos clásicos gracias al trabajo de los traductores que escriben en latín estas obras a partir de los originales griegos o de sus traducciones al árabe. En 1142, Adelardo de Bath (*ca* 1075–1160) tradujo los *Elementos* de Euclides del árabe al latín, traducción que tuvo una gran influencia durante más de un siglo y que dio a conocer en occidente la obra del gran geómetra griego. En 1145, Roberto de Chester, miembro de la Escuela de Traductores de Toledo, tradujo el Almagesto de Ptolomeo del árabe al latín mientras que en 1155, Adelardo hizo una traducción de esta última obra a partir de un original griego. Por este método llegaron también a Europa diversas versiones del *Algebra* de Al-Hwarizmi. Ya en el siglo XIII, Guillermo de Moerbecke, secretario de Santo Tomás de Aquino, publicó una traducción del *Organon* de Aristóteles facilitando así un redescubrimiento de este filósofo para los intelectuales europeos.

Es también a partir de esta época cuando se propaga en Europa el sistema decimal de numeración posicional, que los árabes habían incorporado a su bagaje aritmético desde la India y traído a occidente a través de sus invasiones. Entre los propagadores de este sistema de numeración se encuentra Leonardo de Pisa (ca 1180-1250), más conocido como “Fibonacci”, que era un mercader italiano. Publicó su obra *Liber Abaci* en 1220 que es un tratado sobre métodos y problemas algebraicos en el que recomienda enérgicamente el uso del sistema de numeración posicional de base diez. Escribió también una *Practica Geometriae,* basado en la *Division de las figuras* de Euclides (obra perdida). Es, como su nombre indica, un libro de geometría práctica dedicado a resolver determinados problemas de geometría elemental, sin más pretensiones.

También en el siglo XIII Ramón Llull (1232-1316), principalmente en el *Libre de Contemplació en Déu* (1274-1276), la *Doctrina pueril* (c. 1274) y *el Arbre de ciencia* (1295-1296) y finalmente en su *Libre de Geometria Nova* (1299), presenta una definición de la geometría como una compleja concepción matemático-metafísica del universo. El trasfondo de esta visión geométrica del mundo es pitagórica y platónica; la circunferencia, el cuadrado y el triángulo son la base de toda la realidad y la base de todas las demás figuras geométricas, y también se incluyen mutuamente, formando la “figura plena”, constituida por un círculo, un triángulo y un cuadrado superpuestos y que, al entender de Llull, tienen la misma área, lo que constituye un intento de cuadratura y triangulatura del círculo. Esa imagen representa “la correspondencia analógica entre el mundo divino, el mundo espiritual y el mundo humano.”

Más importante para nuestros objetivos es la figura de Campano de Novara, del que se sabe muy poco de su vida: sólo que fue capellán del papa Urbano VI, cuyo pontificado tuvo lugar entre 1261 y 1264. A él se debe una muy buena traducción del árabe al latín de los Elementos. Estudió los llamados "ángulos corneados", cuestión muy polémica durante la Edad Media y su versión del libro de Euclides ejerció una gran influencia en una obra escrita unos 60 años más tarde y que es el objeto de nuestro estudio en el presente trabajo: la *Geometria Speculativa*, de Thomas Bradwardine (1290? – 1349).

2. Thomas Bradwardine

Thomas Bradwardine forma parte del grupo de filósofos que destacaron en el siglo XIV. No se sabe muy bien ni la fecha ni el lugar de su nacimiento; debió nacer en algún lugar del sur de Inglaterra a finales del siglo XIII, alrededor de 1290.

La primera noticia fidedigna que se tiene de él es su inscripción como "Fellow" del Balliol College de Oxtord en agosto de 1321. Dos años más tarde pasó, también como "Fellow", al Merton College, en el que permaneció presumiblemente hasta 1335. Durante su estancia en este último escribió sus principales obras sobre lógica, matemáticas y filosofía natural.

Se tiene una evidencia explícita que así fue para una de ellas, *Insolubilia*, escrita cuando era "Regent Master in Arts" y hay una gran probabilidad de que escribiera una *Arithmetica Speculativa*, una *Geometria Speculativa*, y seguramente un tratado elemental de astronomía. Ciertamente, tanto la *Arithmetica* como la *Geometria* parecen ser anteriores al *Tractatus de proportionibus*, escrito en 1328, obra que a su vez precede al *Tractatus de continuo*, en la que hace un uso constante de los argumentos geométricos para mejorar la mayoría de puntos de vista medievales sobre la composición de tal concepto.

En la década de los años 1330, Bradwardine estuvo bajo la protección de Richard de Bury, monje benedictino, que había estudiado en Oxford y tenía un gran amor a los libros; fue obispo de Durham entre 1333 y 1345. A instancias de éste, Bradwardine entró en la carrera eclesiástica en 1333 y dos años más tarde se desplazó a Durham al amparo de su protector. Allí, además de tener acceso a la extensa biblioteca de Bury entró en contacto con varios intelectuales de su tiempo, que vivían también bajo el mecenazgo del obispo. En 1337, Bradwardine fue nombrado canciller de la catedral de San Pablo aunque siguió manteniendo un estrecho contacto con Richard de Bury hasta la muerte de éste en 1345.

No se sabe con certeza la fecha en la que fue nombrado capellán del rey Eduardo III. En 1346 acompañó a las tropas inglesas en su expedición a Francia y en agosto de 1348 fue nombrado arzobispo de Canterbury por el Papa, pero el rey consiguió que el Papa anulara dicho nombramiento. No obstante, volvió a ser elegido para el cargo en una segunda oportunidad el 4 de junio de 1349 y fue consagrado un mes más tarde, el 10 de julio, en la sede papal de Avignon. Bradwardine regresó inmediatamente a Inglaterra donde, tras apenas un mes de ejercer como arzobispo, falleció víctima de la peste negra el 26 de agosto de 1349.

3. La Geometria Speculativa

Como ya hemos dicho, durante su estancia en Oxford, Bradwardine escribió la mayoría de sus obras científicas. La que tuvo mayor repercusión fue el *Tractatus de proportionibus*, de 1328, pero con anterioridad había publicado sus tratados de aritmética (la *Arithmetica Speculativa*) y de geometría (la *Geometría Speculativa*). Este último libro comienza diciendo:

La geometría está, de alguna manera, supeditada a la aritmética [igual que la retórica lo está a la dialéctica], pues es posterior a ella en orden, y las propiedades de los números son útiles a las magnitudes, en cuya descripción Euclides interpone la aritmética entre la geometría.

Dice también:

La geometría está dividida en dos partes, teórica y práctica, igual que otras ciencias matemáticas. La teórica es la que investiga las propiedades de las magnitudes a través de razonamientos que hacen uso del silogismo... La práctica es la que investiga las medidas de las magnitudes mediante resultados e instrumentos.

Estos párrafos muestran el primer objetivo del autor, que es el de dar una formación geométrica a los estudiantes de las *artes liberales*. No obstante, el alcance del texto va bastante más allá de este propósito. Así, ya en el siglo XV, Fridericus Amann señala que este libro

... contiene todas las demostraciones geométricas que el filósofo [Aristóteles] enuncia a modo de ejemplo tanto en lógica como en filosofía.

Asimismo, Pedro Sánchez Ciruelo, en 1495, en una versión ampliada del original de Bradwardine dice que la obra

... reúne todas las conclusiones geométricas que necesitan las estudiantes de artes y de la filosofía de Aristóteles.

La necesidad de un amplio bagaje matemático para interpretar correctamente a Aristóteles queda puesta de manifiesto con gran énfasis en las opiniones de Roger Bacon y es evidente por sí mismo en sus propios textos. En este sentido, el libro de Bradwardine resulta adecuado para la formación de los estudiantes, si bien es una obra que rebasa este propósito en muchos sentidos, como veremos más adelante.

Una lectura de la obra pone de manifiesto una influencia evidente de los *Elementos* de Euclides, que Bradwardine recoge a partir de la lectura de dicha obra en la versión de Campano de Novara, y en menor medida, la *Medida del círculo*, de Arquímedes y la *Esférica* de Teodosio. Otro libro importante a lo hora de analizar las fuentes de la Geometria es un libro titulado *Ysoperimetrorum*, que era una traducción (posiblemente de Eutocio, en opinión de A.G. Molland) de una introducción anónima del *Almagesto* de Ptolomeo, adaptada de los trabajos de Zenodoro sobre dicha cuestión.

La *Geometria Speculativa* tuvo un gran éxito, como lo evidencia el gran número de copias que se editaron de la misma. Se conservan una veintena de tales copias y, además, en 1495 se imprimió una versión de la misma de Pedro Sánchez Ciruelo, a la que nos hemos referido anteriormente, que sirvió de modelo a sucesivas ediciones. El propio Sánchez Ciruelo usó este trabajo como base de la sección de geometría de su *Cursus quattuor mathematicarum artium liberaliun*, publicado en Alcalá de Henares en 1516, en la que se pueden leer párrafos que son copia literal de la versión original de Bradwardine. Esto mismo se puede apreciar en una obra de 1390, escrita por Wigandus Durheimer cuyo título, extenso y no falto de pedantería, es *Tractatus geometrie data et ccompilate... ex dictis Euclidis et Campani super geometriam Bradwardini.* Asimismo, Nicolás de Cusa utiliza repetidamente resultados de la *Geometria Speculativa* en sus escritos matemáticos. En el siglo XVI, John Major hace una crítica de la parte del libro que se refiere al infinito y ya en el siglo XVII el matemático polaco Jan Brozek estudia la visión de Bradwardine de los polígonos estrellados.

En el siglo XVI, Juan Luis Vives había criticado con dureza el libro de nuestro autor, que considera correcto, pero

... su estudio no es necesario para los estudiantes, cuyos conocimientos matemáticos son útiles solamente para comprender otras ramas del conocimiento.

Esto da una idea del papel que jugaban en esa época libros como la *Geometria Speculativa* y, en general, el estudio de las matemáticas.

4. Los Postulados

Como hemos dicho en la Introducción, el libro está escrito baja la influencia directa de la versión de Campano de Novara de los *Elementos* de Euclides. Así, la primera de las cuatro partes de que consta el libro comienza, tras una breve introducción, con algunas definiciones para enunciar a continuación los postulados que van a regir a lo largo de la obra y de los que se deducen la proposiciones que siguen, igual que lo hacen Euclides y Campano. Los postulados que presenta Bradwardine difieren de los de estos dos últimos autores en algunos aspectos:

Postulado 1. Trazar una línea recta de cualquier punto a cualquier otro punto. Y ésta es la menor de todas las líneas con tales extremos.
Postulado 2. Desde cualquier punto y ocupando cualquier cantidad de espacio, describir un círculo.
Postulado 3. Todos los ángulos rectos son iguales entre sí.
Postulado 4. Si una línea recta corta otras dos líneas rectas y los ángulos interiores son menores que dos rectos, estas líneas prolongadas en la misma dirección llegan a cortarse.
Postulado 5. Dos líneas rectas no encierran ninguna superficie.

Una simple mirada a estos postulados pone en evidencia algunas diferencias importantes respecto a los de Euclides. En el postulado 1 se añade la frase:

Y ésta [la recta] es la menor de todas las líneas con tales extremos.

Esta frase aparece también en la versión de Campano. La razón de haberla añadido está en la ambigua definición de recta que aparece en los *Elementos*:

Definición 4. Una línea recta es aquella que yace por igual respecto de los puntos que están en ella.

Esta definición de Euclides ha dado lugar a no pocos equívocos y a innumerables debates. Con posterioridad, Arquímedes da la definición de recta en los términos en los que aparece en la adición tanto de Campano como de Bradwardine como la menor de las líneas entre dos puntos. Obsérvese que, por ejemplo, además de la idea intuitiva de recta que todos poseemos, en tanto que línea situada en un plano, si la superficie de referencia es una superficie esférica, la menor línea de la misma entre dos puntos es un arco de la circunferencia del círculo máximo que los contiene.

Vemos también que ha desaparecido el segundo postulado de Euclides, que sí se incluye en la versión de Campano:

Postulado 2. Postúlese prolongar continuamente una recta finita en línea recta.

Esta omisión no es casual, ya que el postulado está en la versión de Campano, y está probada la influencia directa de éste en la obra de Bradwardine. En mi opinión, está relacionada con el hecho de que una recta finita, según Euclides, se puede prolongar sin límites aunque siga siendo finita, lo cual podría chocar con la concepción medieval de un universo geocéntrico finito limitado por la llamada esfera de las estrellas fijas. La ausencia del segundo postulado de los *Elementos* pone también en cuestión el cuarto postulado de la obra que estamos comentando, que requiere precisamente la posibilidad de prolongación de cualquier recta. Este postulado 4, como es bien sabido, está relacionado con la noción de rectas paralelas. Seguramente, para evitar problemas, Bradwardine no da la definición de rectas paralelas pero utiliza este concepto en la proposición 3 sin hacer referencia al significado de este concepto. Esta proposición queda enunciada así:

Si una tercera línea corta dos líneas paralelas, da lugar al mismo tipo y tamaño de ángulos tanto en una como en la otra.

Y antes de dar una demostración de la misma establece tres conclusiones de la misma referidas a los ángulos que se forman cuando dos rectas paralelas se cortan por una recta secante:

1. *Cada ángulo exterior es igual al ángulo interior opuesto.*
2. *Pares de ángulos alternos son iguales.*
3. *Los dos ángulos interiores que se forman a un mismo lado son iguales o suman dos rectos.*

Para evitar problemas de consistencia, se podía haber dado como definición de rectas paralelas tanto al enunciado de la proposición como cualquiera de estas conclusiones o bien prescindir del concepto dando como postulado, simplemente, que los tres ángulos de un triángulo suman dos rectos.

5. La cuestión del espacio infinito

Las dos omisiones de Bradwardine, por otra parte, parecen ir dirigidas a establecer una relación entre la geometría y un mundo finito. Se puede decir que el hecho de no incluir ni el axioma de prolongación ni la definición de rectas paralelas son fruto de un intento de dar un punto de vista realista de la geometría, obviando la posibilidad de considerar la posibilidad de la existencia de un espacio infinito. No obstante, su intento no consigue plenamente sus propósitos y por otra parte, crea una gran cantidad de problemas en el desarrollo de su obra. Digamos que, pese a todo, el propio Bradwardine acabó concibiendo la idea de un espacio infinito con posterioridad a la publicación de su *Geometria Speculativa*.

La posibilidad de la existencia de un mundo infinito es una cuestión ampliamente debatida desde los tiempos de Aristóteles, quien afirmaba que el mundo es finito y no había espacio más allá de la esfera del cielo. No obstante, señalaba que

... los geómetras trabajan con un espacio infinito tridimensional

La cuestión estaba en poner de acuerdo las dos posturas y así en la *Física* (III.4) afirma:

Para muchos, una razón particularmente adecuada y plantea la dificultad que se presenta en cualquiera: no sólo el número, sino también las magnitudes matemáticas y todo cuanto está más allá de los cielos se supone infinito ya que no tiene cabida en nuestra mente.

El propio Aristóteles intenta separar la cuestión matemática de la física y concluye (*Física*, III.8):

El tamaño por el cual una magnitud puede existir en potencia, puede existir en la realidad. Luego, como ninguna magnitud sensible es infinita, es imposible superar cualquier magnitud establecida, pues si fuera posible, sería mayor que los cielos.

Y para justificar las matemáticas, afirma (*Física*, III.8 y *De Coelo*, I.2):

Nuestras afirmaciones no sustraen de su ciencia a los matemáticos, al desaprobar la existencia real del infinito en la dirección del crecimiento, en el sentido de lo transversal. De hecho, no necesitan el infinito y no hacen uso de él. Postulan solamente que la recta finita se puede prolongar tan lejos como se quiera. Es posible dividir en la misma razón de la mayor cantidad otra magnitud del tamaño que queramos. En consecuencia, a los efectos de una demostración, no habrá ninguna diferencia para ellos entre tener dicho infinito en lugar de tales magnitudes, cuya existencia queda inscrita en el marco de las magnitudes reales.

Surge entonces un problema: Si se puede prolongar una recta finita tanto como se quiera, ¿hasta donde se puede llegar? ¿Más allá del cielo, o no?

Parece que la respuesta de Aristóteles es no, pero entonces hay problemas con el axioma de las paralelas de Euclides y para el uso de construcciones que permitan probar ciertos teoremas.

En el siglo XII, Alexander Neckam considera el problema de construir un triángulo equilátero de lado igual al diámetro del mundo, lo que llevaría a un mundo más allá de la esfera de los cielos. Y en el siglo XIII, Robert Grosseteste permitía que la imaginación concibiera un espacio infinito, pero esto no era imaginación en sentido estricto, pues el espacio, para él, no era más que la tridimensionalidad de los cuerpos. También en el siglo XIII, Enrique de Gante manifestaba la dificultad de comprender que no había nada más allá de los cielos, en estrecha relación con las facultades imaginativas.

Bradwardine, en su *De Continuo* expone que el tercer axioma de Euclides (construir un círculo con un punto y una recta) podría ser incompatible con la concepción de un mundo finito:

"De facto" ningún cuerpo puede ser más sutil que el fuego ni puede un círculo ser mayor que el mayor de los círculos celestes, pero estas cosas no son imposibles "per se", como es manifiesto para la inteligencia...

Y Euclides supone lo contrario de lo segundo en la demostración de la primera proposición del primer libro de los Elementos, como queda suficientemente claro si alguien quiere construir un triángulo equilátero sobre el diámetro del mundo.
En *De causa Dei*, la misma cuestión queda resumida con un énfasis más teológico:

Debido a la fuerza y el poder infinitos de Dios y con respecto a Él mismo, junto con todas las cosas creadas en virtud de tal poder, digo que es cierto lo que suponen los geómetras: que una línea recta se puede prolongar de manera continua todo lo que se quiera y un círculo se puede describir sobre cualquier centro y ocupando una cantidad de espacio tan grande como se desee, como se deduce claramente del primer libro de los Elementos de Euclides y que los filósofos naturales suponen, que un medio se pueda enrarecer tanto como se quiera y cosas semejantes, así como cualquier cosa compatible con los lógicos, se dice que son posibles "per se", con tal de que no conlleven contradicción, pese a que no puedan ser posibles por naturaleza, es decir, por un poder natural.

Entendidas así las cosas, los axiomas dela geometría alcanzan una justificación teológica, pero la visión realista de la geometría que tiene Bradwardine hace que se tome muy en serio las consecuencias ontológicas de todo esto.

Esto es particularmente claro en el pasaje del De causa Dei en el que, como consecuencia de la inmutabilidad de Dios, afirma:

Necesariamente, Dios está esencial e inmediatamente en todas partes, no sólo en el mundo y todas sus partes, sino también fuera del mundo, en el imaginable "espacio" infinito o vacío.

Entiende Bradwardine que Dios está claramente en el mundo. Si el “espacio” que ocupa ahora el mundo es A, y B es un “espacio” fuera de mundo, Dios podría mover el mundo de A a B estando entonces en B, pero si no hubiera previamente estado en B se habría movido, lo cual es absurdo. Luego Dios está en todas partes, dentro o fuera del mundo, lo que para Bradwardine es un signo de perfección infinita.

Todo ello da una idea de la idea que tenía Bradwardine de la existencia de un espacio infinito en que desarrollar la geometría.

6. Algunos temas tratados en la obra

Tras los postulados, Bradwardine enuncia unas nociones comunes, al estilo de las que propone Euclides en los *Elementos* y comienza a enumerar una serie de proposiciones en las que trata muchas de las propiedades que aparcen en la versión de Campano del tratado de Euclides y añade algunas cuestiones que no aparecen en el mismo.

6.1. Recubrimiento del plano

La proposición 7 de la primera parte del libro hace referencia al recubrimiento del plano mediante polígonos. No aparece en el original de Euclides, pero es una cuestión tratada en el libro V de la Colección Matemática de Pappus de Alejandría al estudiar las áreas de los polígonos regulares isoperimétricos. Concretamente, en dicho libro aparece esta proposición con la misma redacción que en la *Geometria Speculativa*:

Tres figuras regulares, y no otras, que son el triángulo, el tetrágono y el hexágono, son las que recubren el plano.

6.2. Polígonos estrellados

La primera parte del libro finaliza con varias proposiciones sobre una cuestión escasamente estudiada hasta entonces: los polígonos estrellados o de ángulos emergentes. Dice Bradwardine al abordar este estudio:

Hablaré acerca de figuras con ángulos emergentes de acuerdo con una visión universal y en general, y así, espero que sea suficiente. Pues el tratamiento de este tipo de polígonos es raro y no he encontrado una discusión acerca de ellos, salvo solamente en el caso de Campano, que habla de pasada sobre el pentágono.

Se dice que una figura tiene ángulos emergentes cuando los lados de alguna figura poligonal, tomada entre las más sencillas, se prolongan hasta cortarse en el exterior.

Y nos da las primeras proposiciones sobre este tema:

1. *De entre las figuras con ángulos emergentes, el pentágono es el primero.*
2. *El pentágono de ángulos emergentes tiene cinco ángulos que suman dos rectos.*
3. *Entre las figuras de ángulos emergentes cada sucesor añade dos ángulos rectos a su predecesor.*

Considera a continuación lo que llama polígonos de ángulos emergentes de segundo orden, obtenidos a partir de una segunda intersección de los lados de un polígono regular prolongados y concluye (proposición 4):

De las figuras con ángulos emergente de segundo orden, la primera es el heptágono.

Estudia los polígonos estrellados de órdenes sucesivos y llega a la conclusión expresada en la proposición 5 relativa a esta cuestión:

Al considerar polígonos con ángulos emergentes de cualquier orden, el primero del orden que sucede a otro es el que ocupa el tercer lugar en este último.

Por ejemplo, como el pentágono es el primer polígono estrellado de primer oren, el que ocupa el tercer lugar en el mismo es el heptágono, que es el primero de segundo orden. Asimismo, como en el tercer lugar de los polígonos estrellados de segundo orden es el eneágono, éste será el que genere el primero en el orden tres, etc. Con esta proposición termina la primera parte del libro.

6.3. Ángulos de contingencia

Los ángulos de contingencia es otra de las cuestiones que Bradwardine toma en consideración en su *Geometría.* Comienza dando una definición del concepto:

El ángulo que forma una circunferencia con una línea que la toca recibe el nombre de ángulo de contingencia.

Los ángulos de contingencia, junto con los ángulos curvilíneos los trata Euclides en su *Catóptrica*; fueron objeto de discusión, igual que los ángulos curvilíneos, en la Edad Media por parte de los escolásticos a causa de las peculiaridades de este tipo de ángulos, al no poder ser objeto de medida. Esto queda reflejado en las proposiciones seis y siete del libro de Bradwardine:

Proposición 6. El ángulo de contingencia es menor que cualquier ángulo rectilíneo y es infinitamente divisible.

Bradwardine pasa por alto que la división de un ángulo de contingencia da lugar a dos ángulos de diferente naturaleza, pues uno de ellos es un ángulo de contingencia, mientras el resto forma un ángulo curvilíneo, limitado por dos arcos de circunferencia. Como consecuencia de esta proposición se obtiene:

Un ángulo de contingencia es tanto mayor cuanto menor es el círculo y recíprocamente.

Esta última afirmación se puede ver en La docta ignorancia, de Nicolás de Cusa cuando afirma que la recta es límite de circunferencias o bien que una recta es la circunferencia de radio infinito.

Consecuencia de esta proposición es que un ángulo de contingencia es una magnitud que no admite medida por cuanto es evidente que no es nulo, pero, en cambio, es menor que cualquier ángulo rectilíneo, por pequeño que sea. Es decir, en caso de tener una medida, ésta debería ser menor que cualquier cantidad sin ser cero.

6.4. Figuras isoperimétricas

La segunda parte del libro, finaliza con el estudio de las figuras isoperimétricas, cuestión que no fue considerada por Euclides y sobre la que existen frecuentes referencias tanto en la literatura antigua como en la medieval. Por ejemplo, en astronomía, un argumento estándar que avalaba la esfericidad de la Tierra era que la esfera era el mayor cuerpo sólido entre los que tenían la misma área. Para SACROBOSCO éste era el argumento:

... porque de todos los cuerpos isoperimétricos la esfera es la mayor y de todas las formas, la redondeada es la de mayor capacidad. Por tanto, como el mundo contiene todas las cosas, esta forma era de utilidad y adecuada a él.

Nicolás de Cusa, en el siglo XV, utiliza el método de las figuras isoperimétricas en sus escritos matemáticos para intentar probar la cuadratura del círculo. Pese a la posibilidad de acceder con relativa facilidad a las obras de Zenodoro en los tiempos de Bradwardine, el tratamiento que éste da al tema que nos ocupa está relacionado solo de manera tangencial con el del primero. Tras afirmar que "isoperimétrico" es un término relativo pasa a enunciar su primera conclusión:

Luego, la primera conclusión acerca de los isoperimétricos es esta: Dos figuras son isoperimétricas, una respecto de la otra, si sus perímetros son iguales.

En la *Geometria Speculativa* se afirma:

De todos los polígonos isométricos el que tiene más ángulos es el mayor.

Para la demostración de esta proposición y las dos siguientes se construyen simplemente dos polígonos isoperimétricos, uno de los cuales es mayor que el otro. La proposición siguiente es:

De todos los los polígonos isoperimétricos con el mismo número de ángulos el mayor es el equiangular.

A la que sigue:

De todos los polígonos isoperimétricos con el mismo número de lados y ángulos iguales, el equilátero es el mayor.

Las tres proposiciones representan tres tendencias que el círculo cumple en el más alto grado. Primero, como dice Aristóteles, el círculo es todo ángulo en el mayor grado; en segundo lugar, es siempre igual en todas sus curvas, y tercero, sus lados son todos iguales en el mayor grado puesto que un polígono regular inscrito en el mismo corta su borde en arcos iguales. Así, el círculo cumple perfectamente las tres condiciones anteriores, por lo que se puede concluir que es el de mayor capacidad entre las figuras planas, de la misma manera que la esfera lo es entre los sólidos.

6.5. Razones

La tercera parte del libro se refiere básicamente a las razones. Como sabemos, las razones entre números fueron tratadas en la antigüedad en las tradiciones aritméticas y musicales. El descubrimiento de la inconmensurabilidad de magnitudes geométricas hizo que la teoría de razones de este tipo de magnitudes diera un giro importante. Fue Eudoxo de Cnido quien dio por primera vez la definición de proporcionalidad entre magnitudes geométricas; su teoría está expuesta en el libro V de los Elementos de Euclides. Bradwardine se basa, en principio en la versión de Campano de esta última obra; comienza observando que la geometría tiene un rango más amplio de razones que la aritmética e intenta dar una definición universal de este concepto:

Una razón es un criterio de comparación entre algunas cosas mutuamente comparables entre sí.

El concepto es indeterminado y en realidad, Bradwardine elige un camino distinto al del libro V de Euclides y clasifica las razones en racionales e irracionales, cuestión que en los Elementos no aparece hasta el libro X.

Algunas cantidades se pueden relacionar y decimos que son conmensurables; otras no pueden relacionarse y las llamamos inconmensurables.

Las cantidades conmensurables son aquellas tales que existe una cantidad común que puede numerarlas. Se dice que una cantidad numera a otra si cuando se toma un cierto número de veces da la segunda cantidad, como la línea de un pie da lugar a otra de dos pies y otra de tres. Luego, un línea de dos pies y otra de tres pies pueden relacionarse, pues existe la línea de un pie que las numera como dos y tres.

Pero a las cantidades tales que no existe una cantidad común que las numere se les da el nombre de inconmensurables. De esta clase son la diagonal y el lado de un cuadrado.

Por todo esto, pues, existen dos especies de razones, que son las racionales y las irracionales. Las racionales son las que corresponden a cantidades conmensurables y son las que pueden expresarse mediante números. Una razón irracional se refiere a cantidades inconmensurables, sin que se puedan expresar con números.

En consecuencia, es obvio que el tratamiento de las razones es una cuestión propia de la geometría, pues en este caso toda razón se refiere a magnitudes aunque no todas sean expresables mediante números.

Bradwardine está más interesado en la posibilidad de una descripción numérica de las razones que en saber cómo se podría expresar con exactitud dicha razón. En particular, se plantea como objetivo la posibilidad de establecer una correspondencia biunívoca entre razones y números, y argumenta:

Una razón racional queda denominada inmediatamente por un determinado número, pues al tratarse de cantidades conmensurables es necesario que la menor o una parte de ella numere la mayor conforme a algún número tal como dice Euclides, que afirma *que para todo par de cantidades conmensurables, la razón de una a la otra es como la razón de un número a otro número.*

Con esto, se equiparan las razones entre magnitudes conmensurables a las razones entre números de la aritmética.

Bradwardine esboza la clasificación de las razones racionales siguiendo los criterios de la *Atithmetica* de Boecio de mayor a menor desigualdad y cada una de ellas en cinco especies. Por ejemplo, para las desigualdades mayores, las cinco especies son las razones múltiples (ej. $(3 : 1)$), superparticulares $(4 : 3)$, superpacientes $(7 : 5)$, superparticulares múltiples $(11 : 5)$ y superpacientes múltiples $(11 : 3)$. No identifica estas razones con números o fracciones ni contempla la denominación de una razón como la fracción correspondiente.

En cuanto a las fracciones irracionales, afirma:

Una razón irracional no es inmediatamente denominada por algún número o alguna razón numérica ya que no es posible que alguna parte de la menor numere la mayor según un número. Pero ocurre que una razón irracional se puede denominar de manera intermedia mediante un número.
Por ejemplo, la razón de la diagonal de un cuadrado al lado es la mitad de una razón doble y así, otras especies de estas razones reciben denominaciones numéricas.

Quiere decir el autor que como el cuadrado de lado igual a la diagonal es igual al doble del cuadrado original, la razón doble de la diagonal al cuadrado es de $(1 : 2)$

$$d^2 : l^2 = 1 : 2.$$

No queda claro en la *Geometria* si Bradwardine considera o no que todas las razones irracionales son susceptibles de una denominación numérica. Aborda esta y otras cuestiones en el *Tractatus de Proportionibus* si llegar a definirse. Con un mayor concimiento de la denominación, Nicolás de Oresme pensaba, años más tarde, que era probable que existieran razones que no se pudieran expresar de esta manera.

7. Concusiones

La lectura de la *Geometria Speculativa* sugiere que esta obra es un tratado que pertenece más a la tradición filosófica que a la tradición matemática técnica, especialmente relacionado con problemas del aristotelismo.

Queda patente que Bradwardine intentó adaptar su geometría a un mundo finito pero una lectura de su obra sugiere la existencia de un espacio infinito, para lo que utiliza las formas escolásticas de razonamiento, que afectan de una manera importante el contenido de la obra si bien su implicación en el rigor geométrico es menor que en la tradición matemática griega. Por otra parte, extiende el lenguaje geométrico al estudio de las razones.

La *Geometria Speculativa* fue disminuyendo su influencia a partir del siglo XVI pero fue importante a la hora de inclinar el interés de los filósofos naturales hacia una visión más realista de las matemáticas.

References

[1] T. Bradwardine: *Geometria Speculativa.* Texto original en latín; traducción al inglés y notas de G. Molland. Franz Steiner Verlag, Stutgart 1989.

[2] T. Bradwardine: *Traité des Rapports entre les rapidités dans les mouvements.* Introducción, traducción y comentarios de Sabine Rommevaux. Sagesses médievales, Les Belles Lettres, Paris 2010.

[3] Euclides: *Elementos.* Introducción de Luis Vega; traducción y notas de María Luisa Puertas Castaños. Biblioteca Clásica Gredos, 155, 191 y 228. Editorial Gredos, Madrid 1991, 1994 y 1996.

[4] G. Molland: *An examination of Bradwardine's Geometry.* Archiv. for the Hist. of Exact Sciences **19** (2) (1978) 113 –175.

[5] J. Murdoch: *Thomas Bradwardine.* En *Dictionary of Scientific Biography,* New York (1970–1990).